Docket No.:PC-0039 US

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Commissioner for Patents, Washington, D.C. 20231
on August 4, 2003
By: Diane Kizer
Printed: DIANE KIZER

#14
60
8/14/03

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Chen et al.

Title: MUCIN-RELATED TUMOR MARKER
Serial No.: 09/840,746 Filing Date: April 23, 2001

Examiner: Davis, M.T.B. Group Art Unit: 1642

Mail Stop Appeal Brief-Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION OF DR. TOD BEDILION UNDER 37 C.F.R. § 1.132

I, TOD BEDILION, a citizen of the United States, residing at 132 Winding Way, San Carlos, California, declare that:

1. I was employed by Incyte Corporation (hereinafter "Incyte") as a Director of Corporate Development until May 11, 2001. I am currently under contract to be a Consultant to Incyte Corporation.

2. In 1996, I received a Ph.D. degree in Cell, Molecular and Development Biology from UCLA. I had previously received, in 1988, a B.S. degree in biology from UCLA.

Upon my graduation from UCLA, I became, in April 1996, the first employee of Synteni, Inc. (hereinafter "Synteni"). I was a Research Director at Synteni from April 1996 until Synteni was acquired by Incyte in early 1998.

I understand that Synteni was founded in 1994 by T. Dari Shalon while he was a graduate student at Stanford University. I further understand that Synteni was founded for the purpose of commercially exploiting certain "cDNA microarray" technology that was being worked on at Stanford in the early to mid-1990s. That technology, which I will sometimes refer to herein as the "Stanford-developed cDNA microarray technology", was the subject of Dr. Shalon's doctoral thesis at Stanford. I understand and believe that Dr. P.O. Brown was Dr. Shalon's thesis advisor at Stanford.

During the period beginning before I was employed by Synteni and ending upon its acquisition by Incyte in early 1998, I understand Synteni was the exclusive licensee of the Stanford-developed cDNA microarray technology, subject to any right that the United States government may have with respect to that technology. In early 1998, I understand Incyte acquired rights under the Stanford-developed cDNA microarray technology as part of its acquisition of Synteni.

I understand that at the time of the commencement of my employment at Synteni in April 1996, Synteni's rights with respect to the Stanford-developed cDNA technology included rights under a United States patent application that had been filed June 7, 1995 in the names of Drs. Brown and Shalon and that subsequently issued as United States Patent No. 5,807,522 (the Brown '522 patent). In December 1995, the subject matter of the Brown '522 patent was published based on a PCT patent application that had also been filed in June 1995. The Brown '522 patent (and its corresponding PCT application) describes the use of the Stanford-developed cDNA technology in a number of gene expression monitoring applications, as will be discussed more fully below.

Upon Incyte's acquisition of Synteni, I became employed by Incyte. From early 1998 until late 1999, I was an Associate Research Director at Incyte. In late 1999, I was promoted to the position of Director, Corporate Development.

I have been aware of the Stanford-developed cDNA microarray technology since shortly before I commenced my employment at Synteni. While I was employed by Synteni, virtually all (if not all) of my work efforts (as well as the work efforts of others employed by Synteni) were directed to the further development and commercial exploitation of that cDNA microarray technology. By the end of 1997, those efforts had progressed to the point that I understand Incyte agreed to pay at least about $80 million to acquire Synteni. Since I have been employed by Incyte, I have continued to work on the further development and commercial exploitation of the cDNA microarray technology that was first developed at Stanford in the early to mid-1990s.

3. I have reviewed the specification of a United States patent application that I understand was filed on April 23, 2001 in the names of Huei-Mei Chen et al. and was assigned Serial No. 09/870,746 (hereinafter "the Chen '746 application"). In broad overview, the Chen '746 specification pertains to certain nucleotide and amino acid sequences and their use in a number of applications, including gene expression monitoring applications that are useful in connection with (a) developing drugs (e.g., the diagnosis of inherited and acquired genetic disorders, expression profiling, toxicology testing, and drug development with respect to breast

cancer) and (b) monitoring the activity of drugs for purposes relating to evaluating their efficacy and toxicity.

4. I understand that (a) the Chen '746 application contains claims that are directed to isolated and purified polynucleotides having the sequences of SEQ ID NO:1-encoding polynucleotides, for example SEQ ID NO:2 (hereinafter "the SEQ ID NO:1-encoding polynucleotides"), and (b) the Patent Examiner has rejected those claims on the grounds that the specification of the Chen '746 application does not disclose a substantial, specific and credible utility for the claimed SEQ ID NO:1-encoding polynucleotides. I further understand that whether or not a patent specification discloses a substantial, specific and credible utility for its claimed subject matter is properly determined from the perspective of a person skilled in the art to which the specification pertains at the time of the patent application was filed. In addition, I understand that a substantial, specific and credible utility under the patent laws must be a "real-world" utility.

5. I have been asked (a) to consider with a view to reaching a conclusion (or conclusions) as to whether or not I agree with the Patent Examiner's position that the Chen '746 application does not disclose a substantial, specific and credible "real-world" utility for the claimed SEQ ID NO:1-encoding polynucleotides, and (b) to state and explain the bases for any conclusions I reach. I have been informed that, in connection with my considerations, I should determine whether or not a person skilled in the art to which the Chen '746 application pertains on April 23, 2001 would have concluded that the Chen '746 application disclosed, for the benefit of the public, a specific beneficial use of the SEQ ID NO:1-encoding polynucleotides in their then available and disclosed form. I have also been informed that, with respect to the "real-world" utility requirement, the Patent and Trademark Office instructs its Patent Examiners in Section 2107 of the Manual of Patent Examining Procedure, under the heading "I. Specific and Substantial Requirement," sub-heading "Research Tools":

> "Many research tools such as gas chromatographs, screening assays, and nucleotide sequencing techniques have a clear, specific and unquestionable utility (e.g., they are useful in analyzing compounds). An assessment that focuses on whether an invention is useful only in a research setting thus does not address whether the specific invention is in fact 'useful' in a patent sense. Instead, Office personnel must distinguish between inventions that have a specifically identified substantial utility and inventions whose asserted utility requires further research to identify or reasonably confirm."

6. I have considered the matters set forth in paragraph 5 of this Declaration and have concluded that, contrary to the position I understand the Patent Examiner has taken, the specification of the Chen '746 patent application disclosed to a person skilled in the art at the time of its filing a number of substantial, specific and credible real-world utilities for the claimed SEQ ID NO:1-encoding polynucleotides. More specifically, persons skilled in the art on April 23, 2001 would have understood the Chen '746 application to disclose the use of the SEQ ID NO:1-encoding polynucleotides in a number of gene expression monitoring applications that were well-known at that time to be useful in connection with the development of drugs and the monitoring of the activity of such drugs. I explain the bases for reaching my conclusion in this regard in paragraphs 7-16 below.

7. In reaching the conclusion stated in paragraph 6 of this Declaration, I considered (a) the specification of the Chen '746 application, and (b) a number of published articles and patent documents that evidence gene expression monitoring techniques that were well-known before the April 23, 2001 filing date of the Chen '746 application. The published articles and patent documents I considered are:

(a) Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O., and Davis, R.W., Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes, Proc. Natl. Acad. Sci. USA, 93, 10614-10619 (1996) (hereinafter "the Schena 1996 article") (copy annexed at Tab A);

(b) Schena, M., Shalon, D., Davis, R.W., Brown, P.O., Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray, Science, 270, 467-470 (1995) (hereinafter "the Schena 1995 article") (copy annexed at Tab B);

(c) Shalon and Brown PCT patent application WO 95/35505 titled "Method and Apparatus For Fabricating Microarrays Of Biological Samples," filed on June 16, 1995, and published on December 28, 1995 (hereinafter "the Shalon PCT application") (copy annexed at Tab C);

(d) Brown and Shalon U.S. Patent No. 5,807,522, corresponding to the Shalon PCT application, titled "Methods For Fabricating Microarrays Of Biological Samples," filed on June 7, 1995 and issued on September 15, 1998 (hereinafter "the Brown '522 patent") (copy annexed at Tab D);

(e) DeRisi, J., Penland, L., and Brown, P.O. (Group 1); Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A., and Trent, J.M. (Group 2), Use of a cDNA microarray to analyse gene expression patterns in human cancer, Nat. Genet., 14(4), 457-460 (1996) (hereinafter "the DeRisi article") (copy annexed at Tab E);

(f) Shalon, D., Smith, S.J., and Brown, P.O., A DNA Microarray System for Analyzing Complex DNA Samples Using Two-color Fluorescent Probe Hybridization, Genome Res., 6(7), 639-645 (1996) (hereinafter "the Shalon article") (copy annexed at Tab F);

(g) Heller, R.A., Schena, M., Chai A., Shalon, D., Bedilion, T., Gilmore, J., Woolley, D.E., and Davis R.W., Discovery and analysis of inflammatory disease-related genes using cDNA microarrays, Proc. Natl. Acad. Sci. USA, 94, 2150-2155 (1997) (hereinafter "the Heller article")(copy annexed at Tab G);

(h) Sambrook, J., Fritsch, E.F., Maniatis, T., Molecular Cloning, A Laboratory Manual, pages 7.37 and 7.38, Cold Spring Harbor Press (1989) (hereinafter "the Sambrook Manual") (copy annexed at Tab H);

8. Many of the published articles and patent documents I considered (i.e., at least items (a)-(f) identified in paragraph 7) relate to work done at Stanford University in the early and mid-1990s with respect to the development of cDNA microarrays for use in gene expression monitoring applications under which Synteni became exclusively licensed. As I will discuss, a person skilled in the art who read the Chen '746 application on April 23, 2001 would have understood that application to disclose the SEQ ID NO:1-encoding polynucleotides to be useful for a number of gene expression monitoring applications, e.g., as a probe for the expression of that specific polynucleotide in cDNA microarrays of the type first developed at Stanford.

Furthermore, items (a)-(g) establish that gene expression monitoring applications utilizing cDNA microarrays were well-known and established methods routinely used in toxicology testing and drug development at the time of filing the Chen '746 application and for several years prior to April 23, 2001. As such, one of ordinary skill in the art would have recognized that the SEQ ID NO:1-encoding polynucleotides could be used in toxicology testing and drug development, irrespective of the biochemical activities of the encoded polypeptide.

9. Turning more specifically to the Chen '746 specification, the SEQ ID NO:2 polynucleotide is shown at pp. 3-7 as one of twenty sequences under the heading "Sequence Listing." The Chen '746 specification specifically teaches that the "the invention provides an isolated cDNA ... selected from the group consisting of SEQ ID NO:2...(Chen '746 application at p. 3). It further teaches that (a) the identity of the SEQ ID NO:2 polynucleotide was determined from a human fetal lung tissue library (LUNGFET05) (Chen '746 application, p. 9), (b) the SEQ ID NO:2 polynucleotide encodes for the Mucin-Related Tumor Marker (MRTM

shown as SEQ ID NO:1 (Chen '746 application at p. 3, lines 14-21, and (c) microarray experiments show the differential expression of MRTM in human breast adenocarcinoma cells (BT20) compared with normal breast epithelial cells (Chen '746 application at p. 10, lines 5-8.

The Chen '746 application discusses a number of uses of the SEQ ID NO:1-encoding polynucleotides in addition to their use in gene expression monitoring applications. I have not fully evaluated these additional uses in connection with the preparation of this Declaration and do not express any views in this Declaration regarding whether or not the Chen '746 specification discloses these additional uses to be substantial, specific and credible real-world utilities of the SEQ ID NO:1-encoding polynucleotides. Consequently, my discussion in this Declaration concerning the Chen '746 application focuses on the portions of the application that relate to the use of the SEQ ID NO:1-encoding polynucleotides in gene expression monitoring applications.

10. The Chen '746 application discloses that the polynucleotide sequences disclosed therein, including the SEQ ID NO:1-encoding polynucleotides, are useful as probes in microarrays. It further teaches that the microarrays can be used "to monitor the expression level of large numbers of genes simultaneously" for a number of purposes, including "to develop and monitor the activities of therapeutic agents" (Chen '746 application at p. 14, lines 4-9).

In the same paragraph of the Chen '746 application described above, the Chen '746 application teaches that microarrays can be prepared using the previously mentioned cDNA microarray technology developed at Stanford in the early to mid-1990s. In this connection, the Chen '746 application specifically cites to the Schena 1996 article identified in item (a) of paragraph 7 of this Declaration (Chen '746 application at p. 14 *supra*, paragraph 7).

The Schena 1996 article is one of a number of documents that were published prior to the April 23, 2001 filing date of the Chen '746 application that describes the use of the Stanford-developed cDNA technology in a wide range of gene expression monitoring applications, including monitoring and analyzing gene expression patterns in human cancer. In view of the Chen '746 application, the Schena 1996 article, and other related pre-April 2001 publications, persons skilled in the art on April 23, 2001 clearly would have understood the Chen '746 application to disclose the SEQ ID NO:1-encoding polynucleotides to be useful in cDNA microarrays for the development of new drugs and monitoring the activities of drugs for such purposes as evaluating their efficacy and toxicity, as explained more fully in paragraph 15 below.

With specific reference to toxicity evaluations, those of skill in the art who were working on drug development in April 2001 (and for many years prior to April 2001) without any doubt appreciated that the toxicity (or lack of toxicity) of any proposed drug they were

working on was one of the most important criteria to be considered and evaluated in connection with the development of the drug. They would have understood at that time that good drugs are not only potent, they are specific. This means that they have strong effects on a specific biological target and minimal effects on all other biological targets. Ascertaining that a candidate drug affects its intended target, and identification of undesirable secondary effects (i.e., toxic side effects), had been for many years among the main challenges in developing new drugs. The ability to determine which genes are positively affected by a given drug, coupled with the ability to quickly and at the earliest time possible in the drug development process identify drugs that are likely to be toxic because of their undesirable secondary effects, have enormous value in improving the efficiency of the drug discovery process, and are an important and essential part of the development of any new drug. Accordingly, the teachings in the Chen '746 application, in particular regarding use of the SEQ ID NO:1-encoding polynucleotides in differential gene expression analysis and in the development and the monitoring of the activities of drugs, clearly includes toxicity studies and persons skilled in the art who read the Chen '746 application on April 23, 2001would have understood that to be so.

11. The Schena 1996 article was not the first publication that described the use of the cDNA microarray technique developed at Stanford to monitor quantitatively gene expression patterns. More than a year earlier (i.e., in October 1995), the Schena 1995 article, titled "Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray", was published (see Tabs A and B).

12. As previously discussed (*supra*, paragraphs 2 and 7), in the mid-1990s patent applications were filed in the names of Drs. Shalon and Brown that described the Stanford-developed cDNA microarray technology. The two patent documents (i.e., the Shalon PCT application and the Brown '522 patent) annexed to this Declaration at Tabs C and D evidence information that was available to the public regarding the Stanford-developed cDNA microarray technology before the April 23, 2001 filing date of the Chen '746 application.

The Shalon PCT patent application, which was published in December 1995, contains virtually the same (if not exactly the same) specification as the Brown '522 patent. Hence, the Brown '522 patent disclosure was, in effect, available to the public as of the December 1995 publication date of the Shalon PCT application (see Tabs C and D). For the sake of convenience, I cite to and discuss the Brown '522 specification below on the understanding that the descriptions in that specification were published as of the December 28, 1995 publication date of the Shalon PCT application.

The Brown '522 patent discusses, in detail, the utility of the Stanford-developed cDNA microarrays in gene expression monitoring applications. For example, in the "Summary Of The Invention" section, the Brown '522 patent teaches (see Tab D, col. 4, line 52-col. 5, line 8):

> Also forming part of the invention is a method of detecting differential expression of each of a plurality of genes in a first cell type, with respect to expression of the same genes in a second cell type. In practicing the method, there is first produced fluorescent-labeled cDNAs from mRNAs isolated from two cells types, where the cDNAs from the first and second cell types are labeled with first and second different flourescent reporters.
>
> A mixture of the labeled cDNAs from the two cell types is added to an array of polynucleotides representing a plurality of known genes derived from the two cell types, under conditions that result in hybridization of the cDNAs to complementary-sequence polynucleotides in the array. The array is then examined by fluorescence under fluorescence excitation conditions in which (i) polynucleotides in the array that are hybridized predominantly to cDNAs derived from one of the first or second cell types give a distinct first and second fluorescence emission color, respectively, and (ii) polynucleotides in the array that are hybridized to substantially equal numbers of cDNAs derived from the first and second cell types give a distinct combined fluorescence emission color, respectively. The relative expression of known genes in the two cell types can then be determined by the observed fluorescence emission color of each spot.

The Brown '522 patent further teaches that the "[m]icroarrays of immobilized nucleic acid sequences prepared in accordance with the invention" can be used in "numerous" genetic applications, including "monitoring of gene expression" applications (see Tab D at col. 14, lines 36-42). The Brown '522 patent teaches (a) monitoring gene expression (i) in different tissue types, (ii) in different disease states, and (iii) in response to different drugs, and (b) that arrays disclosed therein may be used in toxicology studies (see Tab D at col. 15, lines 13-18 and 52-58 and col. 18, lines 25-30).

13. Also pertinent to my considerations underlying this Declaration is the DeRisi article, published in December 1996. The DeRisi article describes the use of the Stanford-developed cDNA microarray technology "to analyze gene expression patterns in human cancer" (see Tab E at, e.g., p. 457). The DeRisi article specifically indicates, consistent with what was apparent to persons skilled in the art in December 1996, that increasing the number of

genes on the cDNA microarray permits a "more comprehensive survey of gene expression patterns," thereby enhancing the ability of the cDNA microarray to provide "new and useful insights into human biology and a deeper understanding of the gene pathways involved in the pathogenesis of cancer and other diseases" (see Tab E at p. 458).

14. Other pre-April 2001 publications further evidence the utility of the cDNA microarrays first developed at Stanford in a wide range of gene expression monitoring applications (see, e.g., the Shalon and the Heller articles at Tabs F and G). By no later than the March 1997 publication of the Heller article, these publications showed that employees of Synteni (i.e., James Gilmore and myself) had used the cDNA microarrays in specific gene expression monitoring applications (see Tab G).

The Heller article states that the results reported therein "successfully demonstrate the use of the cDNA microarray system as a general approach for dissecting human diseases" (Tab G at p. 2150). Among other things, the Heller article describes the investigation of "1000 human genes that were randomly selected from a peripheral human blood cell library" and "[t]heir differential and quantitative expression analysis in cells of the joint tissue. . . to demonstrate the utility of the microarray method to analyze complex diseases by their pattern of gene expression" (see Tab G at pp. 2150 *et seq.*).

Much of the work reported on in the Heller article was done in 1996. That article, therefore, evidences how persons skilled in the art were readily able, well prior to April 23, 2001, to make and use cDNA microarrays to achieve highly useful results. For example, as reported in the Heller article, a cDNA microarray that was used in some of the highly successful work reported on therein was made from 1,000 genes randomly selected from a human blood cell library.

15. A person skilled in the art on April 23, 2001, who read the Chen '746 application, would understand that application to disclose the SEQ ID NO:1-encoding polynucleotides, for example, SEQ ID NO:2, to be highly useful as probes for the expression of that specific polynucleotide in cDNA microarrays of the type first developed at Stanford. For example, the specification of the Chen '746 application would have led a person skilled in the art in April 2001 who was using gene expression monitoring in connection with working on developing new drugs for the treatment of breast cancer to conclude that a cDNA microarray that contained the SEQ ID NO:1-encoding polynucleotides would be a highly useful tool and to request specifically that any cDNA microarray that was being used for such purposes contain the SEQ ID NO:1-encoding polynucleotides. Persons skilled in the art would appreciate that cDNA

microarrays that contained the SEQ ID NO:1-encoding polynucleotides would be a more useful tool than cDNA microarrays that did not contain the polynucleotides in connection with conducting gene expression monitoring studies on proposed (or actual) drugs for treating breast cancer for such purposes as evaluating their efficacy and toxicity.

I discuss in more detail in items (a)-(f) below a number of reasons why a person skilled in the art, who read the Chen '746 specification in April 2001, would have concluded based on that specification and the state of the art at that time, that the SEQ ID NO:1-encoding polynucleotides would be a highly useful tool for inclusion in cDNA microarrays for evaluating the efficacy and toxicity of proposed drugs for treating breast cancer, as well as for other evaluations:

(a) The Chen '746 application teaches the SEQ ID NO:1-encoding polynucleotides to be useful as probes in cDNA microarrays of the type first developed at Stanford. It also teaches that such cDNA microarrays are useful in a number of gene expression monitoring applications, including "developing and monitoring the activity of therapeutic agents [i.e., drugs]" (see paragraph 10, *supra*).

(b) By April 2001, the Stanford-developed cDNA microarray technology was a well known and widely accepted tool for use in a wide range of gene expression monitoring applications. This is evidenced, for example, by numerous publications describing the use of that cDNA technology in gene expression monitoring applications and the fact that, for over a year, the technology had provided the basis for the operations of an up-and-running company (Synteni), with employees, that was created for the purpose of developing and commercially exploiting that technology (see paragraphs 2, 8 and 10-14, *supra*). The fact that Incyte agreed to purchase Synteni in late 1997 for an amount reported to be at least about $80 million only serves to underscore the substantial practical and commercial significance, in 1997, of the cDNA microarray technology first developed at Stanford (see paragraph 2, *supra*).

(c) The pre-April 2001 publications regarding the cDNA microarray technology first developed at Stanford that I discuss in this Declaration repeatedly confirm that, consistent with the teachings in the Chen '746 application, cDNA microarrays are highly useful tools for conducting gene expression monitoring applications with respect to the development of drugs and the monitoring of their activity. Among other things, those pre-April 2001 publications confirmed that cDNA microarrays (i) were useful for monitoring gene expression responses to different drugs (see paragraph 12, *supra*), (ii) were useful in analyzing gene expression patterns in human cancer, with increasing the number of genes on the cDNA microarray enhancing the ability of the cDNA microarray to provide useful information (see paragraph 13, supra), and (iii) were a valuable tool for use as part of a "general approach for

dissecting human diseases" and for "analyz[ing] complex diseases by their pattern of gene expression" (see paragraph 14, supra).

(d) Based on my own extensive work for a company whose business was the development and commercial exploitation of cDNA microarray technology for more than two years prior to the April 2001 filing date of the Chen '746 application, I have first-hand knowledge concerning the state of the art with respect to making and using cDNA microarrays as of April 23, 2001 (see paragraphs 2 and 14, *supra)*. Persons skilled in the art as of that date would have (a) concluded that the Chen '746 application disclosed cDNA microarrays containing the SEQ ID NO:1-encoding polynucleotides to be useful, and (b) readily been able to make and use such microarrays with useful results.

(e) The Chen '746 specification contains a number of teachings that would lead persons skilled in the art on April 23, 2001 to conclude that a cDNA microarray that contained the SEQ ID NO:1-encoding polynucleotides would be a more useful tool for gene expression monitoring applications relating to drugs for treating breast cancer than a cDNA microarray that did not contain the SEQ ID NO:1-encoding polynucleotides. Among other things, the Chen '746 specification teaches that the identity of the SEQ ID NO:2 polynucleotide was determined from a fetal lung tissue cDNA library (LUNGFET05) (Chen '746 application, p. 9, lines 25-26). Moreover, microarray experiments show the differential expression of MRTM in human breast adenocarcinoma cells (BT20) compared with normal breast epithelial cells (Chen '746 application at p. 10, lines 5-8. (See paragraph 9, *supra*).

(f) Persons skilled in the art on April 23, 2001 would have appreciated (i) that the gene expression monitoring results obtained using a cDNA microarray containing a probe to a sequence selected from the group consisting of SEQ ID NO:1-encoding polynucleotides would vary, depending on the particular drug being evaluated, and (ii) that such varying results would occur both with respect to the results obtained from the probe described in (i) and from the cDNA microarray as a whole (including all its other individual probes). These kinds of varying results, depending on the identity of the drug being tested, in no way detracts from my conclusion that persons skilled in the art on April 23, 2001, having read the Chen '746 specification, would specifically request that any cDNA microarray that was being used for conducting gene expression monitoring studies on drugs for treating breast cancer (*e.g.*, a toxicology study or any efficacy study of the type that typically takes place in connection with the development of a drug) contain any one of the SEQ ID NO:1-encoding polynucleotides as a probe. Persons skilled in the art on April 23, 2001 would have wanted their cDNA microarray to have a probe as described in (i) because a microarray that contained such a probe (as compared to one that did not) would provide more useful results in the kind of gene expression monitoring

studies using cDNA microarrays that persons skilled in the art have been doing since well prior to April 23, 2001.

The foregoing is not intended to be an all-inclusive explanation of all my reasons for reaching the conclusions stated in this paragraph 15, and in paragraph 6, *supra*. In my view, however, it provides more than sufficient reasons to justify my conclusions stated in paragraph 6 of this Declaration regarding the Chen '746 application disclosing to persons skilled in the art at the time of its filing substantial, specific and credible real-world utilities for the SEQ ID NO:1-encoding polynucleotides.

16. Also pertinent to my considerations underlying this Declaration is the fact that the Chen '746 disclosure regarding the uses of the SEQ ID NO:2 polynucleotide for gene expression monitoring applications is not limited to the use of that polynucleotide as a probe in microarrays. For one thing, the Chen '746 specification teaches that the polynucleotides described therein (including the polynucleotide of SEQ ID NO:2) may desirably be used as probes in any of a number of long established "standard" non-microarray techniques, such as Northern analysis, for conducting gene expression monitoring studies. See, e.g.:

(a) Chen '746 application at p. 8, lines 13-16 ("Probe" refers to a cDNA that hybridizes to at least one nucleic acid in a sample. Where targets are single stranded, probes are complementary single strands. Probes can be labeled with reporter molecules for use in hybridization reactions including Southern, northern, in situ, dot blot, array, and like technologies ...."); and

(b) Chen '746 application at p. 18, lines 11-15 ("In order to provide standards for establishing differential expression, normal and disease expression profiles are established. This is accomplished by combining a sample taken from normal subjects, . . . with a cDNA under conditions for hybridization to occur. Standard hybridization complexes may be quantified by comparing the values obtained using normal subjects with values from an experiment in which a known amount of a purified sequence is used") (emphasis supplied).

The "Sambrook et al." reference is a reference that was well known to persons skilled in the art in August 2000 and is cited in the Chen '746 application at p. 14. A copy of pages from that reference manual, which was published in 1989, is annexed to this Declaration at Tab H. The attached pages from the Sambrook manual provide an overview of northern analysis and other membrane-based technologies for conducting gene expression monitoring studies that were known and used by persons skilled in the art for many years prior to the April 23, 2001 filing date of the Chen '746 application.

A person skilled in the art on April 23, 2001, who read the Chen '746 specification, would have routinely and readily appreciated that the SEQ ID NO:1-encoding polynucleotides disclosed therein would be useful as a probe to conduct gene expression monitoring analyses using northern analysis or any of the other traditional membrane-based gene expression monitoring techniques that were known and in common use many years prior to the filing of the Chen '746 application. For example, a person skilled in the art in April 2001 would have routinely and readily appreciated that the SEQ ID NO:1-encoding polynucleotides would be a useful tool in conducting gene expression analyses, using the northern analysis technique, in furtherance of (a) the development of drugs for the treatment of breast cancer, and (b) analyses of the efficacy and toxicity of such drugs.

17. I declare further that all statements made herein of my own knowledge are true and that all statements made herein on information and belief are believed to be true; and further, that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, and that willful false statements may jeopardize the validity of this application and any patent issuing thereon.

Tod Bedilion

Signed at Redwood City, California
this 21st day of July 2003

Docket No.: PC-0039 US
USSN: 09/840,746
Ref. No. 1 of 7

Website Outages: There will be sporadic outages through the remainder of June.

How Many Genes Are in the Human Genome?

- Subject Index
- Alphabetical Index
- Send the url of this page to a friend

News
- What's New
- Genetics Headlines
- Human Genome News
- Meetings Calendar
- Media Guide

Basic Information
- FAQs
- Glossary
- Acronyms
- Links
- Genetics 101
- Publications

About the Project
- What is it?
- Goals
- Progress
- History
- Ethical Issues

So, how many genes are there?

Although the sequence of the 3 billion bases (A's, C's, G's, and T's) that make up the human genome has been determined, the exact number of genes encoded by the genome is still unknown. The most recent predictions estimate around 30,000 genes, much lower than previous estimates of 80,000 to 140,000.

This lower estimate came as a shock to many scientists, because counting genes was viewed as a way of quantifying genetic complexity. With 30,0000 genes, the human gene count would be only one-third greater than that of the simple roundworm *C. elegans* (~20,000 genes).[1] For scientists who support this lower estimate, biological complexity is explained by gene control and expression rather than by number.

Studies since the publication of the draft genome sequence have generated estimates that differ greatly. A study conducted by scientists at Ohio State University suggests between 65,000 and 75,000 human genes.[2] Another

- Benefits
- Genetics 101

Medicine & the New Genetics
- Home
- Gene Testing
- Gene Therapy
- Pharmacogenomics
- Disease Information
- Genetic Counseling

Ethical, Legal, Social Issues
- Home
- Privacy Legislation
- Gene Testing
- Patenting
- Forensics
- Genetically Modified Food
- Behavioral Genetics
- Minorities, Race, Genetics
- Genetics in Courtroom

Education
- Teachers
- Careers
- Students
- Webcasts Audio/Video
- Images
- Videos
- Chromosome Poster
- Presentations
- Genetics 101
- Genética Websites en Español

Research
- Home
- Sequencing
- Instrumentation
- Mapping
- Bioinformatics
- Functional Genomics
- ELSI Research
- Recent Abstracts
- US,Intl. Research Sites
- Funding

Publications
- Human Genome News
- Chromosome Poster
- Primer Molecular Genetics
- To Know Ourselves
- Your Genes, Your Choices
- 2002 DOE Genome Abstracts
- List of All Publications

Search This Site

search

- Contact Us
- Privacy Statement
- Site Stats and Credits

study published in *Cell* in August 2001 predicted a total human gene count of 42,000.[3]

The ongoing debate has generated a gene-count betting pool called "Gene Sweepstake." Each genome scientist participating in the Cold Spring Harbor Laboratory (CSHL) Genome Meeting is eligible to place a bet on a gene number. The winning number will be determined at the May 2003 CSHL Genome Meeting. The bets range from around 27,000 to more than 150,000. For details and updates on Gene Sweepstake, go to the Web site.

It could be years before a truly reliable gene count can be assessed. The reason for so much discrepancy is that these predictions are derived from different computational methods and gene-finding algorithms. Computation alone is simply not enough to generate an accurate gene number. While gene-counting programs can identify patterns and phenomena that scientists have seen before, the programs are unable to recognize new phenomena. Clearly, gene predictions will have to be verified by labor intensive work in the laboratory before the scientific community can reach any real consensus.[4]

Related Web Sites

Build 33 - Release notes for the most current build of the human genome (based on sequence data available April 10, 2003) used by NCBI in its genome browser called Map Viewer. Predicts 26,846 genes. This is the same build used by the University of California, Santa Cruz (UCSC) Human Genome Browser. More statistics for this build also are provided.

- *Homo sapiens* Genome View - Browse Build 33 of the human genome using NCBI's Map Viewer. Clicking on a chromosome will display a chromosome map that provides the total number of genes mapped to that chromosome.

Ensembl Human release 13.31.1 - The most current human genome release available from the European Bioinformatics Institute's human genome browser. This release was built around NCBI's Build 31. Predicts 24,847 genes.

Count of Mapped Genes by Chromosome - See how many genes have been mapped to each chromosome. Provided by the Genome Database.

Related Scientific Literature

Summary of Public Draft Human Genome Sequence
Lander, Eric S., et al., 2001. "Initial Sequencing and Analysis of the Human Genome." *Nature* **409**, 860-921. Available online (March 2003).

Summary of Celera's Draft Human Genome Sequence
Venter, J. Craig, et al., 2001. "The Sequence of the Human Genome." *Science* **291** (5507), Feb 16 2001: 1304-1351. Available online (March 2003).

"The Nature of the Number," 2000 *Nature Genetics* **25**, 127-28. (Editorial). Available online.

Aparicio, Samuel A. J. R., 2000. "How to Count...Human Genes." *Nature Genetics* **25**, 129-30. Available online.

Ewing, B., and Green, P., 2000. "Analysis of Expressed Sequence Tags Indicates 35,000 Human Genes." *Nature Genetics* **25**, 232-34. Available online.

Crollius, Hugues Roest, et al., "Estimate of Human Gene Number Provided by Genome-wide Analysis Using *Tetraodon nigroviridis* DNA Sequence." *Nature Genetics* **25,** 235-38. Available online.

Liang, Feng, et al., "Gene Index Analysis of the Human Genome Estimates Approximately 120,000 Genes." *Nature Genetics* **25**, 239-40. Available online.

References

1. Claverie, Jean-Michel, 2001. "Gene number. What if there are only 30,000 human genes?" *Science* **291**, 1255-7. Available online (March 2003).

2. Briggs, Helen, 2001. "Dispute Over Number of Human Genes." *BBC News Online.* (March 2003).

3. Wright, Fred A., et al., 2001. "A draft annotation and overview of the human genome." *Genome Biology* **2,** 1-18. (March 2003).

4. Hollon, Tom, 2001. "Human Genes: How Many?" *The Scientist* **15**, 1. (March 2003).

Send the url of this page to a friend

Site sponsored by the U.S. Department of Energy Office of Science, Office of Biological and Environmental Research, Human Genome Program

Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 9228–9232, September 1994
Biochemistry

Construction and characterization of a normalized cDNA library

(brain mRNA/DNA circles/reassociation kinetics)

MARCELO BENTO SOARES*†, MARIA DE FATIMA BONALDO*†, PIERRE JELENE*†, LONG SU*†, LEE LAWTON*†, AND ARGIRIS EFSTRATIADIS‡

Departments of *Psychiatry and ‡Genetics and Development, Columbia University, and †The New York State Psychiatric Institute, 722 West 168th Street, New York, NY 10032

Communicated by Isidore S. Edelman, June 8, 1994

ABSTRACT **We have developed a simple procedure based on reassociation kinetics that can reduce effectively the high variation in abundance among the clones of a cDNA library that represent individual mRNA species. For this normalization, we used as a model system a library of human infant brain cDNAs that were cloned directionally into a phagemid vector and, thus, could be easily converted into single-stranded circles. After controlled primer extension to synthesize a short complementary strand on each circular template, melting and reannealing of the partial duplexes at relatively low C_0t, and hydroxyapatite column chromatography, unreassociated circles were recovered from the flow through fraction and electroporated into bacteria, to propagate a normalized library without a requirement for subcloning steps. An evaluation of the extent of normalization has indicated that, from an extreme range of abundance of 4 orders of magnitude in the original library, the frequency of occurrence of any clone examined in the normalized library was brought within the narrow range of only 1 order of magnitude.**

The mRNAs of a typical somatic cell are distributed in three frequency classes (1, 2) that are presumably maintained in representative cDNA libraries. The classes at the two extremes (*ca.* 10% and 40–45% of the total, respectively) include members occurring at vastly different relative frequencies. On average, the most prevalent class consists of about 10 mRNA species, each represented by 5000 copies per cell, whereas the class of high complexity comprises 15,000 different species each represented by 1–15 copies only. Rare mRNAs are even more under represented in the brain, a tissue exhibiting an exceptionally high sequence complexity of transcripts (3–5).

Although even the rarest mRNA sequence from any tissue is likely to be represented in a cDNA library of 10^7 recombinants, its identification is very difficult (its frequency of occurrence may be as low as 2×10^{-6} on average or even 10^{-7} for complex tissues such as the brain). Thus, for a variety of purposes, it is advantageous to apply a normalization procedure and bring the frequency of each clone in a cDNA library within a narrow range (generation of a perfectly equimolar cDNA library is practically impossible in our experience). Normalized cDNA libraries can facilitate positional cloning projects aiming at the identification of disease genes, can increase the efficiency of subtractive hybridization procedures, and can significantly facilitate genomic research pursuing chromosomal assignment of expressed sequences and their localization in large fragments of cloned genomic DNA (exon mapping). Normalization makes feasible the gridding of cDNA libraries on filters at high density by reducing the number of clones to be arrayed (gridding 10^7 clones for 1× coverage of a non-normalized library is not a feasible task). Finally, by increasing the frequency of occurrence of rare cDNA clones while decreasing simultaneously the percentage of abundant cDNAs, normalization can expedite significantly the development of expressed sequence databases by random sequencing of cDNAs.

Although cDNA library normalization could be achieved by saturation hybridization to genomic DNA (6), this approach is impractical, since it would be extremely difficult to provide saturating amounts of the rarer cDNA species to the hybridization reaction. The alternative is the use of reassociation kinetics: assuming that cDNA reannealing follows second-order kinetics, rarer species will anneal less rapidly and the remaining single-stranded fraction of cDNA will become progressively normalized during the course of the reaction (6–8). As we report here, we have used this kinetic principle to develop a method for normalization of a directionally cloned cDNA library that has significant advantages over two previously reported similar procedures (refs. 7 and 8; see *Results and Discussion*).

MATERIALS AND METHODS

cDNA Library Construction. Poly(A)$^+$ RNA isolated from the entire brain of a female infant (72 days old), who died in consequence of spinal muscular atrophy, was used for construction of a cDNA library (IB) as described (9, 10). As a primer for first-strand cDNA synthesis, we used the oligonucleotide 5′-AACTGGAAGAATTCGCGGCCGCAG-GAAT$_{18}$-3′, which contains a *Not* I site (underlined). After ligation to *Hin*dIII adaptors, the cDNAs were digested with *Not* I and cloned directionally into the *Hin*dIII and *Not* I sites of a phagemid vector (L-BA) constructed by modifying pEMBL-9(+) (11). L-BA carries an ampicillin-resistance gene, plasmid and filamentous phage (f1) origins of replication, and cloning sites (5′ *Hin*dIII–*Bam*HI–*Not* I–*Eco*RI 3′). Superinfection of bacteria with the helper phage M13K07 (12) converts duplex plasmids into single-stranded DNA circles containing message-like strands of the cDNA inserts.

Preparation of Single-Stranded Library DNA. Plasmid DNA from the IB library was electroporated into *Escherichia coli* DH5α F′ bacteria, and the culture was grown under ampicillin selection at 37°C to an OD_{600} of 0.2, superinfected with a 20-fold excess of the helper phage M13K07, and harvested after 4 hr for preparation of single-stranded plasmids, as described (12). To eliminate contaminating double-stranded, replicative form (RF) DNA, 20 μg of the preparation was digested with *Pvu*II (which cleaves only duplex DNA molecules), extracted with phenol/chloroform, diluted by addition of 2 ml of loading buffer (0.12 M sodium phosphate buffer, pH 6.8/10 mM EDTA/1% SDS), and purified by hydroxyapatite (HAP) chromatography at 60°C, using a column preequilibrated with the same buffer (1-ml bed volume; 0.4 g of HAP). After a 6-ml wash with loading buffer,

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: HAP, hydroxyapatite.

this volume was combined with the flow through fraction, and the sample was extracted twice with water-saturated 2-butanol, once with dry 2-butanol, and once with water-saturated ether (3 volumes per extraction). The sample was desalted by passage through a Nensorb column (DuPont/NEN) according to the manufacturer's specifications, concentrated by ethanol precipitation, and electrophoresed in a low-melting agarose gel to remove helper phage DNA and any residual tRNA contaminant or oligoribonucleotides (breakdown products from the RNase A digestion used during purification). The region of the gel containing the single-stranded library was excised and, after β-agarase (New England Biolabs) digestion, the DNA was purified and ethanol-precipitated.

cDNA Library Normalization. The IB cDNA library was normalized (see Fig. 1) in two consecutive rounds to derive the normalized libraries ^{1}NIB and ^{2}NIB, by using the following procedure. To synthesize a partial second strand of about 200 nt by limited extension, 9 pmol of the oligonucleotide primer 5′-GGCCGCAGGAAT$_{15}$-3′ was added to 4.5 pmol of single-stranded IB library DNA in a 150-μl reaction mixture containing 30 mM Tris·HCl (pH 7.5); 50 mM NaCl; 15 mM $MgCl_2$; 1 mM dithiothreitol; 0.1 mM dNTPs; 2.5 mM ddATP, ddCTP, and ddGTP; and a trace of [α-^{32}P]dCTP. The mixture was incubated for 5 min at 60°C and for 15 min at 50°C, the temperature was lowered to 37°C, 75 units of Klenow DNA polymerase (United States Biochemical) was added, and the incubation was continued for 30 min. The reaction was terminated by addition of EDTA (20 mM), extracted with phenol/chloroform, diluted with 2 ml of HAP loading buffer containing 50 μg of sonicated and denatured salmon sperm DNA carrier, and chromatographed on HAP, as described above. After washing, the partial duplex circles bound to HAP were eluted from the column with 6 ml of 0.4 M phosphate buffer, pH 6.8/10 mM EDTA/1% SDS. The concentration of phosphate in the eluate was lowered to 0.12 M by addition of 14 ml of water containing 50 μg of DNA carrier, and the chromatographic step was repeated. The final eluate was extracted and desalted as described above and the DNA was ethanol-precipitated. The pellet (112 ng) was dissolved in 2.5 μl of formamide and the sample was heated for 3 min at 80°C under a drop of mineral oil to dissociate the DNA strands. For an annealing reaction, the volume was brought to 5 μl by adding 0.5 μl of 0.1 M Tris·HCl, (pH 7.5) containing 0.1 M EDTA, 0.5 μl of 5 M NaCl, 1 μl (5 μg) of $(dT)_{25-30}$, and 0.5 μl (0.5 μg) of the extension primer. The last two ingredients were added to block stretches of adenine residues [representing the initial poly(A) tails] and regions complementary to the oligonucleotide on the single-stranded DNA circles. The annealing mixture was incubated at 42°C, and a 0.5-μl aliquot was withdrawn at 13 hr (calculated C_0t, 5.5). The unhybridized single-stranded circles (normalized library) were separated from the reassociated partial duplexes by HAP chromatography and then recovered from the flow through fraction as described above. Since we, and others (13), have observed that the electroporation efficiency of partially repaired circular molecules is increased by about 100-fold in comparison with single-stranded circles, the normalized cDNA circles were converted to partial duplexes by primer extension using random hexamers and T7 DNA polymerase (Sequenase version II; United States Biochemical), in a 10–20 μl reaction mixture containing 1 mM dNTPs. After addition of EDTA to 20 mM, phenol extraction, and ethanol precipitation, the cDNAs were dissolved in 10 mM Tris·HCl, pH 7.5/1 mM EDTA, and electroporated into competent bacteria (DH10B; GIBCO/BRL). To determine the number of transformants, 1 hr after the electroporation a 10-μl aliquot of the culture was plated on an LB agar plate containing 75 μg/ml ampicillin (extrapolation from these data indicated that a normalized library of 2.5×10^6 colonies was obtained). Supercoiled plasmid DNA was then prepared (^{1}NIB library) with a Qiagen plasmid kit (Qiagen, Chatsworth, CA). The same protocol was used for a second round of normalization (calculated C_0t, 2.5) to derive the ^{2}NIB library (1.3×10^7 transformants) from a preparation of ^{1}NIB single-stranded circles, except that the HAP purification step after primer extension to synthesize short complementary strands was omitted.

Colony Hybridization. For screening, colonies were grown on duplicate nylon filters (GeneScreen*Plus*; DuPont/NEN) that were processed as described (14) and hybridized at 42°C in 50% formamide/5× Denhardt's solution/0.75 M NaCl/0.15 M Tris·HCl, pH 7.5/0.1 M sodium phosphate/0.1% sodium pyrophosphate/2% SDS containing sheared and denatured salmon sperm DNA at 100 μg/ml. Radioactive probes were prepared by random primed synthesis (15, 16) using the Prime-it II kit (Stratagene).

DNA Sequencing. Double-stranded plasmid DNA templates were prepared by using the Wizard Minipreps DNA purification system (Promega) and sequenced from both ends by using the universal forward and reverse M13 fluorescent primers. Reactions were assembled on a Biomek 1000 workstation (Beckman) and then transferred to a thermocycler (Perkin–Elmer/Cetus) for cycle sequencing. Reaction products were analyzed on an automated 370A DNA sequencer (Applied Biosystems). Nucleic acid and protein database searches were performed at the National Center for Biotechnology Information server using the BLAST algorithm (17).

RESULTS AND DISCUSSION

Experimental Strategy. To develop a normalization procedure, shown schematically in Fig. 1, and at the same time

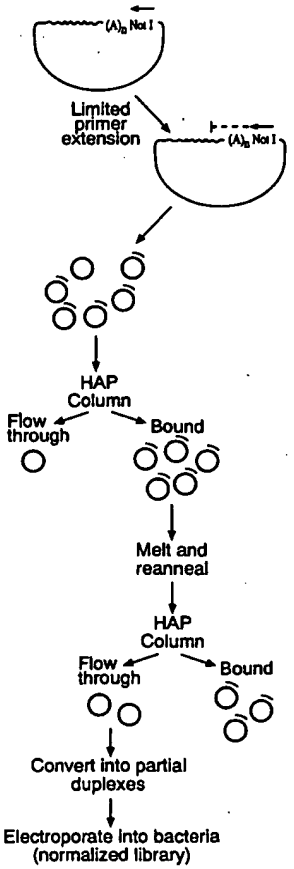


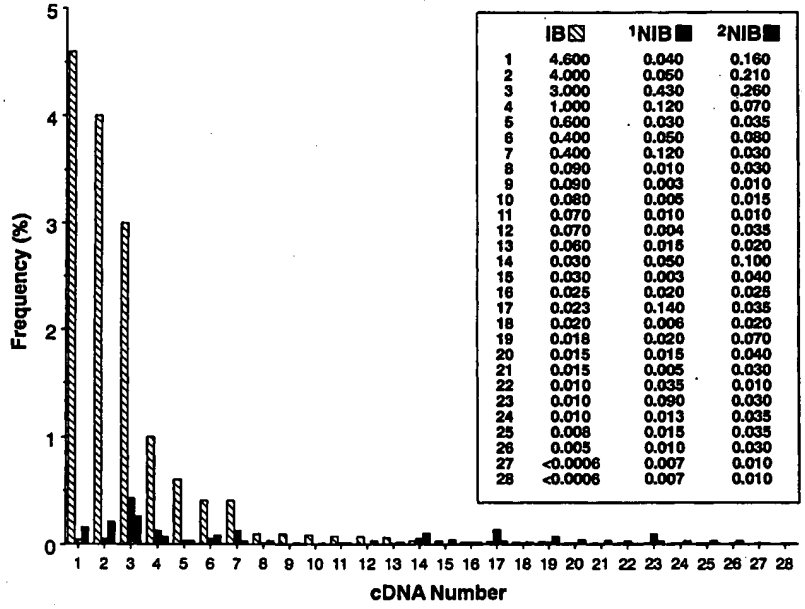
FIG. 1. Diagram of the normalization procedure. Single-stranded circles of a library of directionally cloned cDNAs are primer extended under controlled conditions to generate complementary strands of about 200 ± 20 nt, and the resulting partial duplexes are purified from unprimed circles by HAP chromatography. Bound DNA is melted and reannealed to a relatively low C_0t (see text). The remaining single-stranded circles (normalized library) are isolated by HAP chromatography, converted into partial duplexes by random priming, and electroporated into bacteria for amplification.

increase the utility of the normalized model cDNA library, we first constructed a high-quality brain cDNA library (IB) that has the following features (10): the average size of a cDNA insert is 1.7 kb, often providing coding-region information by sequencing from the 5′ end; the length of the segment representing the mRNA poly(A) tail is short, allowing an increase in the output of useful sequencing information from the 3′ end; the frequency of nonrecombinant clones is extremely low (0.1%); and chimeric cDNAs have not been encountered, after single-pass sequencing of >2000 clones (10, 18). However, the latter analysis also demonstrated that 13% of the clones in the IB library lacked poly(A) tails and were presumably derived from aberrant priming.

To preserve the length of the cDNAs, avoid differential loss of sequences, and alleviate a need for subcloning steps after normalization, we excluded from our protocol the use of PCR and chose directional cloning into a phagemid vector. Such vectors have been previously used advantageously for cDNA library subtractions (13), although normalization was not attempted. This cloning regime readily provides single strands that can be used both for annealing and for direct propagation in bacteria. In control experiments (data not shown), we assessed the frequency of occurrence of abundant cDNAs (encoding α- and β-tubulin, elongation factor 1α, and myelin basic protein) and demonstrated that, at least by this criterion, the representation of clones in the starting library remained unchanged after conversion into single-stranded circles. We also note that electrophoretic purification of the circles prior to use is necessary, to remove contaminating oligoribonucleotides (see *Materials and Methods*), whose presence would result in undesirable internal priming events during the first step of our protocol.

In contrast with our scheme, two other PCR-based normalization methods (7, 8) necessitate the use of subcloning steps. In one of these approaches (7), sheared cDNAs (0.2–0.4 kb) were ligated to a linker–primer, amplified by PCR, normalized kinetically, reamplified, and finally cloned directionally in such a way that only 3′-terminal sequences (almost exclusively 3′ noncoding regions) were purposely preserved. The steps of the second scheme (8) were similar, except that the process started from cloned, randomly primed, and relatively short cDNAs, initially selected to minimize length-dependent differential PCR amplification. Thus, both coding and noncoding regions were represented in the final normalized library, but in pieces.

While maintaining length and representation of mRNA regions, our protocol (Fig. 1) also addresses successfully the problem recognized in the first of the alternative approaches (7). It was considered that the 3′ noncoding region is almost always unique to the transcript that it represents and is expected, therefore, to anneal only to its complement. In contrast, cross-hybridization of coding regions belonging to unequally represented members of oligo- or multigene families could result in the elimination of rarer members from the population during the normalization process. This possibility is precluded in our method, which begins with the synthesis, from the 3′ end of the cDNA, of a short complementary strand on the circular single-stranded cDNA template under controlled conditions, calibrated to yield strands with a narrow size distribution (200 ± 20 nt). Since the average length of 3′ noncoding regions in brain mRNAs is 750 nt (19), the vast majority of synthesized complementary strands participating in the annealing reaction should be devoid of coding region sequences. However, after this partial extension step, purification of the products by HAP chromatog-



	IB	¹NIB	²NIB
1	4.600	0.040	0.160
2	4.000	0.050	0.210
3	3.000	0.430	0.260
4	1.000	0.120	0.070
5	0.600	0.030	0.035
6	0.400	0.050	0.080
7	0.400	0.120	0.030
8	0.090	0.010	0.030
9	0.090	0.003	0.010
10	0.080	0.005	0.015
11	0.070	0.010	0.010
12	0.070	0.004	0.035
13	0.060	0.015	0.020
14	0.030	0.050	0.100
15	0.030	0.003	0.040
16	0.025	0.020	0.025
17	0.023	0.140	0.035
18	0.020	0.006	0.020
19	0.018	0.020	0.070
20	0.015	0.015	0.040
21	0.015	0.005	0.030
22	0.010	0.035	0.010
23	0.010	0.090	0.030
24	0.010	0.013	0.035
25	0.008	0.015	0.035
26	0.005	0.010	0.030
27	<0.0006	0.007	0.010
28	<0.0006	0.007	0.010

FIG. 2. Comparison of the frequencies of cDNA probes in the original (IB) and two normalized (¹NIB and ²NIB) libraries. The indicated percentages of 28 cDNA sequences in the three libraries, tabulated in order of decreasing frequency in the IB library, are shown in the form of a histogram to visualize normalization. Frequencies were calculated from the number of positive colonies after hybridization of duplicate filters containing 500–180,000 colonies from each of the three cDNA libraries with the following 28 probes: 1, elongation factor 1α; 2, α-tubulin; 3, β-tubulin; 4, myelin basic protein; 5, aldolase; 6, 89-kDa heat shock protein; 7, γ-actin; 8, secretogranin; 9, microtubule-associated protein; 11, vimentin; 13, a cDNA randomly picked from the ¹NIB library similar to a mouse cysteine-rich intestinal protein (¹NIB-2, GenBank accession nos. T09996 and T09997); 19, a cDNA isolated from the ¹NIB library homologous to the human endogenous retrovirus RTVLH2 (cDNA-20, accession nos. L13822 and L13823); 20, histone H2b.1; 23, a cDNA randomly picked from the ¹NIB library encoding the human polyposis (*DP1* gene) mRNA (¹NIB-227, accession nos. T10266 and T10267); 27, a cDNA randomly picked from the ¹NIB library related to the human endogenous retrovirus ERV9 gene (¹NIB-114, accession nos. T10086 and T10087); the remaining brain cDNAs are novel, and except for nos. 10, 18, 21, and 25, they were randomly picked from the ¹NIB library.

raphy is necessary to eliminate single strands of the IB library lacking poly(A) tails that cannot participate in primed synthesis. We repeat the chromatographic step to reduce the background to negligible levels, since after the first passage through the HAP column about 0.1% of pure single strands bind nonspecifically. However, during the second round of normalization to derive the ^{2}NIB library, we omitted this step since we showed that 187 clones, which were picked randomly and sequenced from the ^{1}NIB library (see below), all contained 3′ poly(A) stretches. The remaining steps of our procedure entail melting and reannealing of the partial duplexes, followed by purification of unreassociated circles (normalized library) by HAP chromatography and electroporation into bacteria (Fig. 1).

Characterization of Normalized cDNA Libraries. To evaluate the extent of normalization achieved with our method, we compared the IB, ^{1}NIB, and ^{2}NIB libraries by colony hybridization. For this analysis, we used 28 cDNA probes chosen to represent various frequencies of occurrence within a wide range (at least 4 orders of magnitude: 4.6% to <0.0006%) in the IB library (Fig. 2). However, an additional comparison of these results with independent theoretical estimates was necessary, to provide a further assessment of the degree of normalization, especially because the ^{1}NIB library was derived after incubation to a relatively low C_0t (5.5) during the reannealing step of our procedure. When relatively high C_0t values were used in our initial attempts to normalize the IB library, we obtained unsatisfactory results (high background) that we attribute to technical problems inherent to the procedure. Nevertheless, a reevaluation of brain cDNA hybridization data (ref. 20; see Table 1) suggests that a relatively low C_0t would suffice for our purpose, to bring the frequency of each library clone within a narrow range.

For our calculations (Table 1), which should be regarded as rough but indicative estimates, we used a set of reliable hybridization data that are available only for mouse brain mRNAs (20), assuming that these measurements should not differ significantly among mammals (in all cases examined, including humans, the average amount of RNA per brain cell and the number of cells per gram of tissue are practically the same; see, e.g., refs. 29 and 30). These calculations show that at C_0t 5.5, of the three kinetic classes of mRNAs, the most abundant species are drastically diminished, while all frequencies are brought within the range of 1 order of magnitude (Table 1, compare columns b and h and columns f and i). Our experimental results (Fig. 2) show that the same range was achieved after a single round of normalization at this C_0t (5.5). Thus, for all practical purposes, a single cycle is probably sufficient. Secondary normalization (calculated C_0t = 2.5) to derive the ^{2}NIB library, although it did not result in a dramatic improvement, preserved the range of frequencies, while making the differences among individual sequences narrower overall (Fig. 2). Eleven of the 28 probes used in this analysis were derived from clones that were randomly picked from the ^{1}NIB library. The overall frequency fold variation was reduced from >7667 (4.6/<0.0006) in the IB library to 133 (0.4/0.003) and 26 (0.1/0.01) in the ^{1}NIB and ^{2}NIB libraries, respectively. However, some unexplained anomalies were also observed for a small minority of clones, whose already reduced frequencies in the ^{1}NIB library were somewhat increased in the ^{2}NIB library (Fig. 2).

To provide a further indication that normalization was successful, we sequenced from both ends 187 cDNA clones that were randomly picked from the ^{1}NIB library (GenBank accession numbers T09994–T10011 and T10014–T10369). With the exception of 4 clones, which carried sequences corresponding to human mitochondrial 16S rRNA, all other cDNAs of this pool were unique, in agreement with the expectation for a normalized library. To further investigate the effect of the normalization procedure on the subset of mitochondrial 16S rRNA clones (1.4%, 1%, and 0.4% in the IB, ^{1}NIB and ^{2}NIB libraries, respectively), we compared the sequences of a number of 16S rRNA clones isolated from both the IB and ^{1}NIB libraries (kindly provided by M. Adams, Institute for Genomic Research and J. Sikela, University of Colorado). This analysis (data not shown) revealed that the 16S rRNA clones isolated from ^{1}NIB did not correspond to

Table 1. Estimates of frequencies of brain mRNAs

Component[a]	%[b]	k_{pfo} (pure)[c]	Complexity,[d] kb	No. of RNA species[e]	Frequency per species,[f] %	k_{so}[g]	Component at C_0t 5.5,[h] %	Final frequency per species,[i] %
I	16	10	96	36	0.44	6.15	0.7	0.02
II	46	0.165	5,800	2,150	0.02	0.10	44.2	0.02
III	38	0.0079	122,000	45,000	0.0008	0.0048	55.1	0.0012

[a]The experimental data of pseudo-first-order hybridization kinetics of cDNA tracer, which was synthesized from mouse brain poly(A)$^+$ polysomal mRNA and driven by its template (20), were solved by computer (unconstrained fit) into three kinetic components, using the EXCESS function of a least-squares curve-fitting program (21).

[b]The fraction of total occupied by each of the components is shown, after a minor correction (at completion, practically all of the tracer had reacted). These numbers (and all other numbers) in the table have been rounded.

[c]The computer-calculated pseudo-first-order hybridization rate constant (k_{pfo}; M^{-1}·sec^{-1}) for each component was divided by each of the values in column b, to derive k_{pfo} (pure).

[d]The complexity (i.e., length of unique sequence) was calculated by considering the data from a calibration kinetic standard: cDNA synthesized from encephalomyocarditis virus RNA (complexity, 9.7 kb) that was driven by its template [k_{pfo} (pure), 99]. Thus, each of the values in column d is the ratio $(99 \times 9.7)/k_{pfo}$ (pure). The complexity calculated for the rarest component (III) matches closely the values obtained from additional kinetic experiments using cDNA enriched for infrequent sequences (22, 23) and also the data of saturation experiments with single-copy genomic DNA tracer (24, 25).

[e]The number of different RNA species in each component was estimated from their complexities by assuming that the average size of brain mRNA is 2.7 kb (26). A conjecture (26) that rare brain mRNAs are longer than this value (hypothetically 5 kb on average) has not been supported by hard evidence.

[f]The initial average frequency of an individual mRNA species of each component in the entire population of mRNA molecules is the ratio of values in column b to those in column e.

[g]To assess the behavior of these kinetic components under the annealing conditions that we used for normalization (C_0t, 5.5; length of complementary sequence in annealing strands, 0.2 kb), we first calculated the second-order reassociation rate constant (k_{so}; M^{-1}·sec^{-1}) for each component. For this calculation, we considered that the k_{so} of a single and pure kinetic component with a complexity of 1 kb reacting at a fragment length of 0.2 kb is 590 (27, 28). Thus each k_{so} value is 590 divided by the complexity in column d.

[h]To determine the percentage of the leftover of each component in the population at C_0t 5.5, we first used the k_{so} values in column g to calculate the fraction remaining single-stranded, according to the equation $C/C_0 = 1/(1 + kC_0t)$ and then normalized the derived values to a total of 100%.

[i]The final average frequency of an individual mRNA species of each component is the ratio of values in column h to those in column e.

the predominant 16S rRNA species present in the IB library. Interestingly, in 17 of 19 16S rRNA clones sequenced from the IB library, the position of the A tract was the same as that present in the mature 16S rRNA. In contrast, all 8 clones sequenced from the ^{1}NIB library represented truncated versions of the 16S rRNA, in which different lengths of the 3′ terminal sequence were absent. Such truncated clones are under represented in the IB library (2 of 19). Therefore, their frequency was increased by normalization, as expected, while the 16S rRNA clones of the most prevalent form were reduced. It is likely that the shorter clones represent bona fide copies of naturally occurring truncated 16S rRNA molecules (ref. 31–33; to be discussed elsewhere).

Database searches (both BLASTN and BLASTX; ref. 17) revealed that of the 183 cDNAs examined, 152 (83%) were unknown (no hits), 15 (8.2%) corresponded to known human sequences, 5 (2.7%) were novel but related to known human sequences, 4 (2.2%) were homologous to mammalian sequences, and 7 (3.8%) were homologous to known sequences from various nonmammalian organisms.

In contrast to these results, when 1633 randomly picked clones from the non-normalized IB library were sequenced mostly (88%) from the 5′ end, the percentage of unknown sequences was significantly lower than in our case (63%), while about 30% of the clones were sequenced twice or more (up to 50) times (10). Similar results were obtained by sequencing 493 random IB clones exclusively from the 3′ end (18). Of the initially abundant cDNAs, which were sequenced multiple times in both of these studies, those encoding elongation factor 1α, α-tubulin, β-tubulin, myelin basic protein, and γ-actin (corresponding to our probes 1–4 and 7; Fig. 2) were absent from the pool of 187 clones that we examined. Moreover, only 15 of the unique 183 clones that we sequenced from the ^{1}NIB library (8%) had been previously identified in the collection of the sequenced 1633 IB clones.

Eighteen of the unknown cDNAs that we sequenced (10% of the total clones) carried *Alu* repetitive elements (6 at the 5′ end; 11 at the 3′ end; and 1 at both ends). Thus, as previously observed (8), the frequency of cDNAs containing *Alu* repeats is not reduced by normalization. This phenomenon can be attributed to sequence heterogeneity among *Alu* family members, which are able to form imperfect hybrids that probably cannot bind to HAP. However, this is not a disadvantageous property, since it prevents elimination of rare *Alu*-carrying cDNAs from the population.

To assess whether the normalization procedure had skewed the distribution of lengths favoring shorter cDNA clones, Southern blots of released inserts from the IB, ^{1}NIB, and ^{2}NIB plasmids were hybridized with several of the cDNA probes used in Fig. 2 individually. The results (not shown) demonstrated that the intensity of hybridization signals varied as expected, but the size of each hybridizing fragment remained the same.

Note. Sasaki *et al.* (34) have described an alternative normalization procedure, in which a cDNA library was constructed following depletion of abundant mRNA species by sequential cycles of hybridization to matrix-bound cDNA. However, this procedure does not seem to be more advantageous than ours, while its actual practical potential remains to be assessed, as the putative normalized library was not adequately characterized.

We thank Greg Lennon (Lawrence Livermore National Laboratories) for arraying 40,000 clones of the ^{1}NIB library and making it available to other investigators; M. Adams and J. Sikela for kindly providing sequences of 16S rRNA clones isolated from the IB and ^{1}NIB libraries, respectively; M. Adams for making a set of 1633 sequences available to us prior to publication; and Rachel Yarmolinsky for help with the figures. This work was supported by Grant FG02-91ER61233 from the U.S. Department of Energy and by Grants 1R55HD28422 and BRSG 2 S07 RR05395-30 from the National Institutes of Health to M.B.S.; by Grants HG00362, HG00424, and HG00861 from the National Institutes of Health to A.E.; and by a generous gift from the W. M. Keck Foundation.

1. Davidson, E. H. & Britten, R. J. (1979) *Science* **204**, 1052–1059.
2. Bishop, J. O., Morton, J. G., Rosbash, M. & Richardson, M. (1974) *Nature (London)* **250**, 199–204.
3. Hahn, W. E. & Owens, G. P. (1988) in *The Molecular Biology of Neurological Disease*, eds. Rosenberg, R. N. & Harding, A. E. (Butterworths, London), pp. 22–34.
4. Kaplan, B. & Finch, C. (1982) in *Molecular Approaches to Neurobiology*, ed. Brown, I. (Academic, New York), pp. 71–98.
5. Snider, B. J. & Morrison-Bogorad, M. (1992) *Brain Res. Rev.* **17**, 263–282.
6. Weissman, S. M. (1987) *Mol. Biol. Med.* **4**, 133–143.
7. Ko, M. S. H. (1990) *Nucleic Acids Res.* **18**, 5705–5711.
8. Patanjali, S. R., Parimoo, S. & Weissman, S. M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1943–1947.
9. Soares, M. B. (1994) in *Automated DNA Sequencing and Analysis Techniques*, ed. Venter, J. C. (Academic, London), pp. 110–114.
10. Adams, M. D., Soares, M. B., Kerlavage, A. R., Fields, C. & Venter, J. C. (1993) *Nat. Genet.* **4**, 373–380.
11. Dente, L., Cesareni, G. & Cortese, R. (1983) *Nucleic Acids Res.* **11**, 1645–1655.
12. Vieira, J. & Messing, J. (1987) *Methods Enzymol.* **153**, 3–11.
13. Rubenstein, J. L. R., Brice, A. E. J., Ciaranello, R. D., Denney, D., Porteus, M. H. & Usdin, T. B. (1990) *Nucleic Acids Res.* **18**, 4833–4842.
14. Grunstein, M. & Hogness, D. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3961–3965.
15. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
16. Feinberg, A. P. & Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266–267.
17. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
18. Khan, A. S., Wilcox, A. S., Polymeropoulos, M. H., Hopkins, J. A., Stevens, T. J., Robinson, M., Orpana, A. K. & Sikela, J. M. (1992) *Nat. Genet.* **2**, 180–185.
19. Hawkins, J. D. (1988) *Nucleic Acids Res.* **16**, 9893–9908.
20. Hahn, W. E., Van Ness, J. & Maxwell, I. H. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5544–5547.
21. Pearson, W. R., Davidson, E. H. & Britten, R. J. (1977) *Nucleic Acids Res.* **4**, 1727–1737.
22. Van Ness, J. & Hahn, W. E. (1982) *Nucleic Acids Res.* **10**, 8061–8077.
23. Chaudhari, N. & Hahn, W. E. (1983) *Science* **220**, 924–928.
24. Bantle, J. A. & Hahn, W. E. (1976) *Cell* **8**, 139–150.
25. Grouse, L. D., Schrier, B. K., Bennett, E. L., Rosenzweig, M. R. & Nelson, P. G. (1978) *J. Neurochem.* **30**, 191–203.
26. Milner, R. J. & Sutcliffe, J. G. (1983) *Nucleic Acids Res.* **11**, 5497–5520.
27. Galau, G. A., Britten, R. J. & Davidson, E. H. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 1020–1023.
28. Welsh, J., Liu, J.-P. & Efstratiadis, A. (1990) *Genet. Anal. Technol. Appl.* **7**, 5–17.
29. Mandel, P., Rein, H., Harth-Edel, S. & Mardell, R. (1964) in *Comparative Neurochemistry*, ed. Richter, D. (Macmillan, New York), pp. 149–163.
30. Winick, M. (1968) *Pediatr. Res.* **2**, 352–355.
31. Mazo, A. M., Minchenko, A. G., Avdonina, T. A., Gause, G. G. & Pusyriov, A. T. (1983) *Mol. Biol. Rep.* **9**, 155–161.
32. Baserga, S. J., Linnenbach, A. J., Malcolm, S., Ghosh, P., Malcolm, A. D. B., Takeshita, K., Forget, B. G. & Benz, E. J., Jr. (1985) *Gene* **35**, 305–312.
33. Christianson, T. W. & Clayton, D. A. (1988) *Mol. Cell. Biol.* **8**, 4502–4509.
34. Sasaki, Y. F., Ayusawa, D. & Oishi, M. (1994) *Nucleic Acids Res.* **22**, 987–992.

Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 8945–8947, August 1997
Applied Biological Sciences

Docket No.: PC-0039 US
USSN: 09/840,746
Ref. No. 4 of 7

Whole genome analysis: Experimental access to all genome sequenced segments through larger-scale efficient oligonucleotide synthesis and PCR

DEVAL A. LASHKARI*†, JOHN H. MCCUSKER‡, AND RONALD W. DAVIS*§

*Departments of Genetics and Biochemistry, Beckman Center, Stanford University, Stanford, CA 94305; and ‡Department of Microbiology, 3020 Duke University Medical Center, Durham, NC 27710

Contributed by Ronald W. Davis, May 20, 1997

ABSTRACT **The recent ability to sequence whole genomes allows ready access to all genetic material. The approaches outlined here allow automated analysis of sequence for the synthesis of optimal primers in an automated multiplex oligonucleotide synthesizer (AMOS). The efficiency is such that all ORFs for an organism can be amplified by PCR. The resulting amplicons can be used directly in the construction of DNA arrays or can be cloned for a large variety of functional analyses. These tools allow a replacement of single-gene analysis with a highly efficient whole-genome analysis.**

The genome sequencing projects have generated and will continue to generate enormous amounts of sequence data. The genomes of *Saccharomyces cerevisiae*, *Escherichia coli*, *Haemophilus influenzae* (1), *Mycoplasma genitalium* (2), and *Methanococcus jannaschii* (3) have been completely sequenced. Other model organisms have had substantial portions of their genomes sequenced as well, including the nematode *Caenorhabditis elegans* (4) and the small flowering plant *Arabidopsis thaliana* (5). This massive and increasing amount of sequence information allows the development of novel experimental approaches to identify gene function.

One standard use of genome sequence data is to attempt to identify the functions of predicted open reading frames (ORFs) within the genome by comparison to genes of known function. Such a comparative analysis of all ORFs to existing sequence data is fast, simple, and requires no experimentation and is therefore a reasonable first step. While finding sequence homologies/motifs is not a substitute for experimentation, noting the presence of sequence homology and/or sequence motifs can be a useful first step in finding interesting genes, in designing experiments and, in some cases, predicting function. However, this type of analysis is frequently uninformative. For example, over one-half of new ORFs in *S. cerevisiae* have no known function (6). If this is the case in a well studied organism such as yeast, the problem will be even worse in organisms that are less well studied or less manipulable. A large, experimentally determined gene function database would make homology/motif searches much more useful.

Experimental analysis must be performed to thoroughly understand the biological function of a gene product. Scaling up from classical "cottage industry" one-gene-oriented approaches to whole-genome analysis would be very expensive and laborious. It is clear that novel strategies are necessary to efficiently pursue the next phase of the genome projects—whole-genome experimental analysis to explore gene expression, gene product function, and other genome functions. Model organisms, such as *S. cerevisiae*, will be extremely important in the development of novel whole-genome analysis techniques and, subsequently, in improving our understanding of other more complex and less manipulable organisms.

The genome sequence can be systematically used as a tool to understand ORFs, gene product function, and other genome regions. Toward this end, a directed strategy has been developed for exploiting sequence information as a means of providing information about biological function (Fig. 1). Efforts have been directed toward the amplification of each predicted ORF or any other region of the genome ranging from a few base pairs to several kilobase pairs. There are many uses for these amplicons—they can be cloned into standard vectors or specialized expression vectors, or can be cloned into other specialized vectors such as those used for two-hybrid analysis. The amplicons can also be used directly by, for example, arraying onto glass for expression analysis, for DNA binding assays, or for any direct DNA assay (7). As a pilot study, synthetic primers were made on the 96-well automated multiplex oligonucleotide synthesizer (AMOS) instrument (8) (Fig. 2). These oligonucleotides were used to amplify each ORF on yeast chromosome V. The current version of this instrument can synthesize three plates of 96 oligonucleotides each (25 bases) in an 8-hr day. The amplification of the entire set of PCR products was then analyzed by gel electrophoresis (Fig. 3). Successful amplification of the proper length product on the first attempt was 95%. This project demonstrates that one can go directly from sequence information to biological analysis in a truly automated, totally directed manner.

These amplicons can be incorporated directly in arrays or the amplicons can be cloned. If the amplicons are to be cloned, novel sequences can be incorporated at the 5′ end of the oligonucleotide to facilitate cloning. One potential problem with cloning PCR products is that the cloned amplicons may contain sequence alterations that diminish their utility. One option would be to resequence each individual amplicon. However, this is expensive, inefficient, and time consuming. A faster, more cost-effective, and more accurate approach is to apply comparative sequencing by denaturing HPLC (9). This method is capable of detecting a single base change in a 2-kb heteroduplex. Longer amplicons can be analyzed by use of appropriate restriction fragments. If any change is detected in a clone, an alternate clone of the same region can be analyzed. Modifying the system to allow high throughput analysis by denaturing HPLC is also relatively simple and straightforward.

If amplicons are used directly on arrays without cloning, it is important to note that, even if single PCR product bands are observed on gels, the PCR products will be contaminated with various amounts of other sequences. This contamination has the potential to affect the results in, for example, expression

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

†Present address: Synteni, Inc., 6519 Dumbarton Circle, Fremont, CA 94555.
§To whom reprint requests should be addressed at: Department of Biochemistry, Beckman Center, B400, Stanford University, Stanford, CA 94305-5307. e-mail: gilbert@cmgm.stanford.edu.

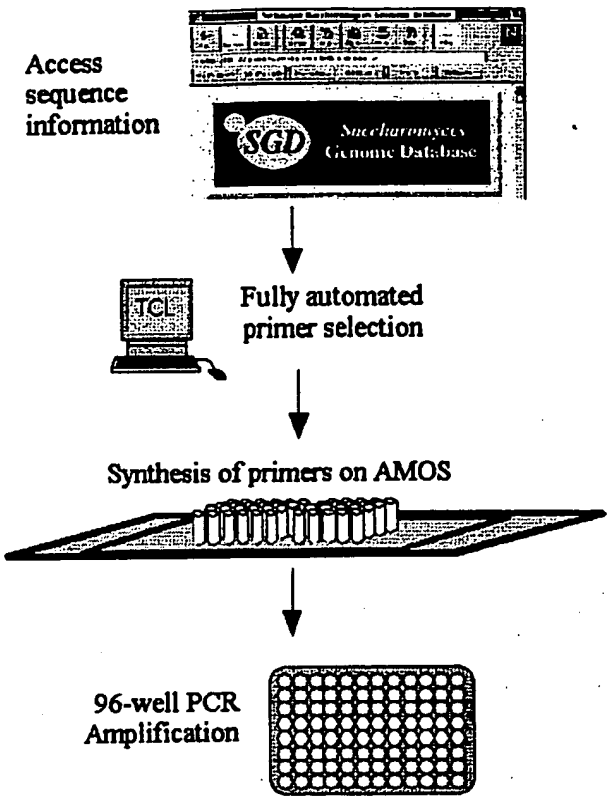


FIG. 1. Overview of systematic method for isolating individual genes. Sequence information is obtained automatically from sequence databases. The data are input into primer selection software specifically designed to target ORFs as designated by database annotations. The output file containing the primer information is directly read by a high-throughput oligonucleotide synthesizer, which makes the oligonucleotides in 96-well plates (AMOS, automated multiplex oligonucleotide synthesizer). The forward and reverse primers are synthesized in the same location on separate plates to facilitate the downstream handling of primers. The amplicons are generated by PCR in 96-well plates as well.

analysis. On the other hand, direct use of the amplicons is much less labor intensive and greatly decreases the occurrence of mistakes in clone identification, a ubiquitous problem associated with large clone set archiving and retrieving.

Any large-scale effort to capture each ORF within a genome must rely on automation if cost is to be minimized while efficiency is maximized. Toward that end, primers targeting ORFs were designed automatically using simple new scripts and existing primer selection software. These script-selected primer sequences were directly read by the high-throughput synthesizer and the forward and reverse primers were synthesized in separate plates in corresponding wells to facilitate automated pipetting and PCR amplifications. Each of the resulting PCR products, generated with minimum labor, contains a known, unique ORF.

Large-scale genome analysis projects are dependent on newly emerging technologies to make the studies practical and economically feasible. For example, the cost of the primers, a significant issue in the past, has been reduced dramatically to make feasible this and other projects that require tens of thousands of oligonucleotides. Other methods of high-throughput analysis are also vital to the success of functional analysis projects, such as microarraying and oligonucleotide chip methods (10–14).

Changes in attitude are also required. One of the major costs of commercial oligonucleotides is extensive quality control such that virtually 100% of the supplied oligonucleotides are successfully synthesized and work for their intended purpose.

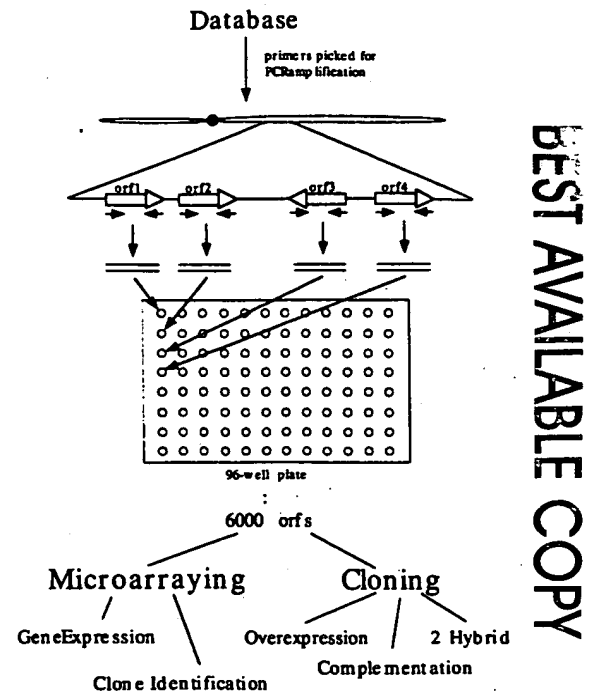


FIG. 2. Overall approach for using database of a genome to direct biological analysis. The synthesis of the 6,000 ORFs (orfs) for each gene of *S. cerevisiae* can be used in many applications utilizing both cloning and microarraying technology.

Considerable cost reduction can be obtained by simply decreasing the expected successful synthesis rate to 95–97%. One can then achieve faster and cheaper whole genome coverage by simply adding a single quality control at the end of the experiment and batching the failures for resynthesis.

The directed nature of the amplicon approach is of clear advantage. The sequence of each ORF is analyzed automatically, and unique specific primers are made to target each ORF. Thus, there is relatively little time or labor involved—for example, no random cloning and subsequent screening is required because each product is known. In the test system, primers for 240 ORFs from chromosome V were systematically synthesized, beginning from the left arm and continuing through to the right arm. At no point was there any manual analysis of sequence information to generate the collection. In many ways, now that the sequence is known, there is no need for the researcher to examine it.

These amplicons can be arrayed and expression analysis can be done on all arrayed ORFs with a single hybridization (10). Those ORFs that display significant differential expression patterns under a given selection are easily identified without the laborious task of searching for and then sequencing a clone. Once scaled up, the procedure provides even greater returns on effort, because a single hybridization will ultimately provide a "snapshot" of the expression of all genes in the yeast genome. Thus, the limiting factor in whole genome analysis will not be the analysis process itself, but will instead be the ability of researchers to design and carry out experimental selections.

Current expression and genetic analysis technologies are geared toward the analysis of single genes and are ill suited to analyze numerous genes under many conditions. Additional difficulties with current technologies include: the effort and expense required to analyze expression and make mutants, the potential duplication of effort if done by different laboratories, and the possibility of conflicting results obtained from different laboratories. In contrast, whole genome analysis not only is more efficient, it also provides data of much higher quality; all genes are assayed and compared in parallel under exactly

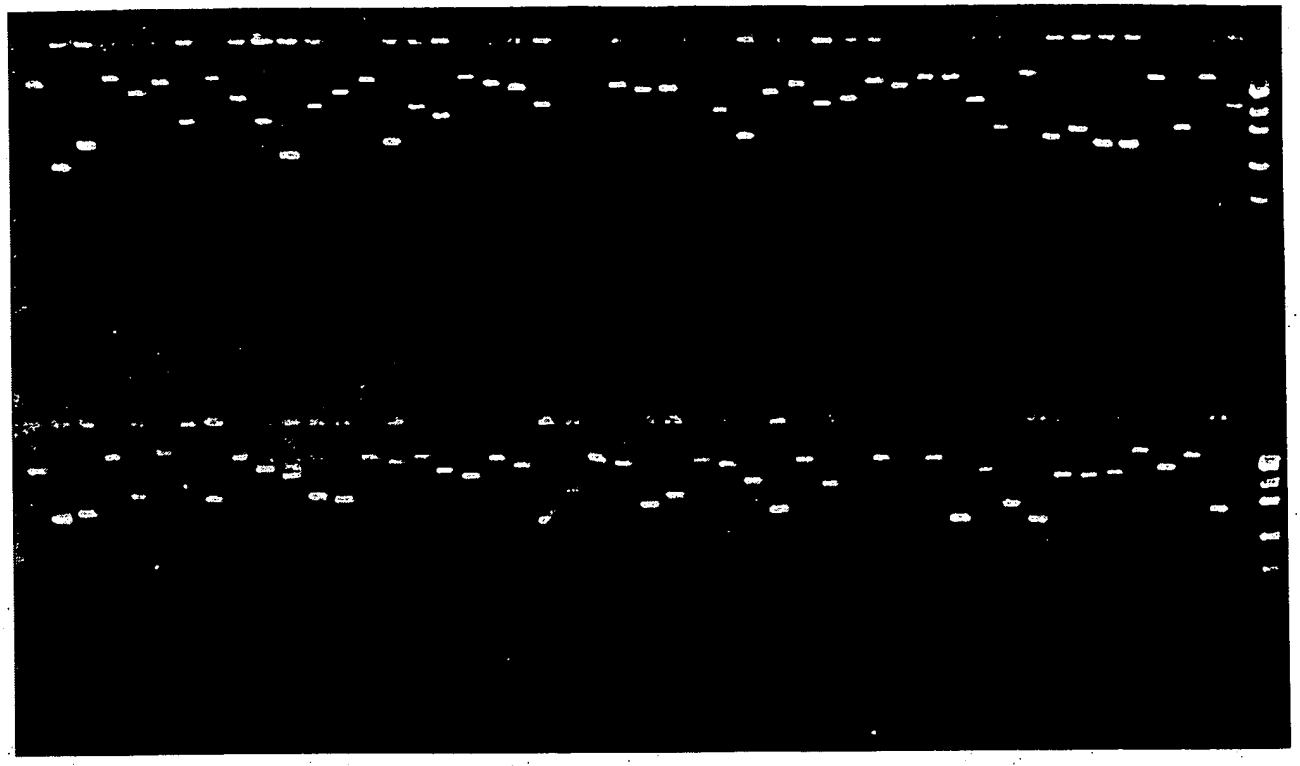


FIG. 3. Gel image of amplifications. Using the method described in Fig. 1, amplicons were generated for ORFs of *S. cerevisiae* chromosome V. One plate of 96 amplification reactions is shown.

the same conditions. In addition, amplicons have many applications beyond gene expression. For example, one recent approach is to incorporate a unique DNA sequence tag, synthesized as part of each gene specific primer, during amplification. The tags or molecular bar codes, when reintroduced into the organism as a gene deletion or as a gene clone, can be used much more efficiently than individual mutations or clones because pools of tagged mutants or transformants can be analyzed in parallel. This parallel analysis is possible because the tags are readily and quantitatively amplified even in complex mixtures of tags (13).

These ORF genome arrays and oligonucleotide tagged libraries can be used for many applications. Any conventional selection applied to a library that gives discrete or multiple products can use these technologies for a simple direct readout. These include screens and selections for mutant complementation, overexpression suppression (15, 16), second-site suppressors, synthetic lethality, drug target overexpression (17), two-hybrid screens (18), genome mismatch scanning (19), or recombination mapping.

The genome projects have provided researchers with a vast amount of information. These data must be used efficiently and systematically to gain a truly comprehensive understanding of gene function and, more broadly, of the entire genome which can then be applied to other organisms. Such global approaches are essential if we are to gain an understanding of the living cell. This understanding should come from the viewpoint of the integration of complex regulatory networks, the individual roles and interactions of thousands of functional gene products, and the effect of environmental changes on both gene regulatory networks and the roles of all gene products. The time has come to switch from the analysis of a single gene to the analysis of the whole genome.

Support was provided by National Institutes of Health Grants R37H60198 and P01H600205.

1. Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., *et al.* (1995) *Science* **269**, 496–512.
2. Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., *et al.* (1995) *Science* **270**, 397–403.
3. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., *et al.* (1996) *Science* **273**, 1058–1073.
4. Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N., Green, P., Thierry-Mieg, J., Qiu, L., Dear, S., Coulson, A., Craxton, M., Durbin, R., Berks, M., Metzstein, M., Hawkins, T., Ainscough, R. & Waterston, R. (1992) *Nature (London)* **356**, 37–41.
5. Newman, T., de Bruijn, F. J., Green, P., Keegstra, K., Kende, H., *et al.* (1994) *Plant Physiol.* **106**, 1241–1255.
6. Oliver, S. (1996) *Nature (London)* **379**, 597–600.
7. Lashkari, D. A. (1996) Ph.D. dissertation (Stanford Univ., Stanford, CA).
8. Lashkari, D. A., Hunicke-Smith, S. P., Norgren, R. M., Davis, R. W. & Brennan, T. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 7912–7915.
9. Oefner, P. J. & Underhill, P. A. (1995) *Am. J. Hum. Genet.* **57**, A266.
10. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. (1995) *Science* **270**, 467–470.
11. Fodor, S. P., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T. & Solas, D. (1991) *Science* **251**, 767–773.
12. Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S. & Fodor, S. P. (1996) *Science* **274**, 610–614.
13. Shoemaker, D. D., Lashkari, D. A., Morris, D., Mittmann, M. & Davis, R. W. (1996) *Nat. Genet.* **14**, 450–456.
14. Smith, V., Chou, K., Lashkari, D., Botstein, D. & Brown, P. O. (1996) *Science* **274**, 2069–2074.
15. Magdolen, V., Drubin, D. G., Mages, G. & Bandlow, W. (1993) *FEBS Lett.* **316**, 41–47.
16. Ramer, S. W., Elledge, S. J. & Davis, R. W. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 11589–11593.
17. Rine, J., Hansen, W., Hardeman, E. & Davis, R. W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 6750–6754.
18. Fields, S. & Song, O. (1989) *Nature (London)* **340**, 245–246.
19. Nelson, S. F., McCusker, J. H., Sander, M. A., Kee, Y., Modrich, P. & Brown, P. O. (1994) *Nat. Genet.* **4**, 11–18.

Docket No.: PC-0039 US
USSN: 09/840,746
Ref. No. 6 of 7

Application of DNA Arrays to Toxicology

John C. Rockett and David J. Dix

Reproductive Toxicology Division, National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, USA

DNA array technology makes it possible to rapidly genotype individuals or quantify the expression of thousands of genes on a single filter or glass slide, and holds enormous potential in toxicologic applications. This potential led to a U.S. Environmental Protection Agency-sponsored workshop titled "Application of Microarrays to Toxicology" on 7–8 January 1999 in Research Triangle Park, North Carolina. In addition to providing state-of-the-art information on the application of DNA or gene microarrays, the workshop catalyzed the formation of several collaborations, committees, and user's groups throughout the Research Triangle Park area and beyond. Potential application of microarrays to toxicologic research and risk assessment include genome-wide expression analyses to identify gene-expression networks and toxicant-specific signatures that can be used to define mode of action, for exposure assessment, and for environmental monitoring. Arrays may also prove useful for monitoring genetic variability and its relationship to toxicant susceptibility in human populations. ***Key words:*** **DNA arrays, gene arrays, microarrays, toxicology.** ***Environ Health Perspect*** **107:681–685 (1999). [Online 6 July 1999]**
http://ehpnet1.niehs.nih.gov/docs/1999/107p681-685rockett/abstract.html

Decoding the genetic blueprint is a dream that offers manifold returns in terms of understanding how organisms develop and function in an often hostile environment. With the rapid advances in molecular biology over the last 30 years, the dream has come a step closer to reality. Molecular biologists now have the ability to elucidate the composition of any genome. Indeed, almost 20 genomes have already been sequenced and more than 60 are currently under way. Foremost among these is the Human Genome Mapping Project. However, the genomes of a number of commonly used laboratory species are also under intensive investigation, including yeast, *Arabidopsis*, maize, rice, zebra fish, mouse, rat, and dog. It is widely expected that the completion of such programs will facilitate the development of many powerful new techniques and approaches to diagnosing and treating genetically and environmentally induced diseases which afflict mankind. However, the vast amount of data being generated by genome mapping will require new high-throughput technologies to investigate the function of the millions of new genes that are being reported. Among the most widely heralded of the new functional genomics technologies are DNA arrays, which represent perhaps the most anticipated new molecular biology technique since polymerase chain reaction (PCR).

Arrays enable the study of literally thousands of genes in a single experiment. The potential importance of arrays is enormous and has been highlighted by the recent publication of an entire *Nature Genetics* supplement dedicated to the technology (*1*). Despite this huge surge of interest, DNA arrays are still little used and largely unproven, as demonstrated by the high ratio of review and press articles to actual data papers. Even so, the potential they offer has driven venture capitalists into a frenzy of investment and many new companies are springing up to claim a share of this rapidly developing market.

The U.S. Environmental Protection Agency (EPA) is interested in applying DNA array technology to ongoing toxicologic studies. To learn more about the current state of the technology, the Reproductive Toxicology Division (RTD) of the National Health and Environmental Effects Research Laboratory (NHEERL; Research Triangle Park, NC) hosted a workshop on "Application of Microarrays to Toxicology" on 7–8 January 1999 in Research Triangle Park, North Carolina. The workshop was organized by David Dix, Robert Kavlock, and John Rockett of the RTD/NHEERL. Twenty-two intramural and extramural scientists from government, academia, and industry shared information, data, and opinions on the current and future applications for this exciting new technology. The workshop had more than 150 attendees, including researchers, students, and administrators from the EPA, the National Institute of Environmental Health Sciences (NIEHS), and a number of other establishments from Research Triangle Park and beyond. Presentations ranged from the technology behind array production through the sharing of actual experimental data and projections on the future importance and applications of arrays. The information contained in the workshop presentations should provide aid and insight into arrays in general and their application to toxicology in particular.

Array Elements

In the context of molecular biology, the word "array" is normally used to refer to a series of DNA or protein elements firmly attached in a regular pattern to some kind of supportive medium. DNA array is often used interchangeably with gene array or microarray. Although not formally defined, microarray is generally used to describe the higher density arrays typically printed on glass chips. The DNA elements that make up DNA arrays can be oligonucleotides, partial gene sequences, or full-length cDNAs. Companies offering pre-made arrays that contain less than full-length clones normally use regions of the genes which are specific to that gene to prevent false positives arising through cross-hybridization. Sequence verification of cDNA clone identity is necessary because of errors in identifying specific clones from cDNA libraries and databases. Premade DNA arrays printed on membranes are currently or imminently available for human, mouse, and rat. In most cases they contain DNA sequences representing several thousand different sequence clusters or genes as delineated through the National Center for Biotechnology Information UniGene Project (*2*). Many of these different UniGene clusters (putative genes) are represented only by expressed sequence tags (ESTs).

Array Printing

Arrays are typically printed on one of two types of support matrix. Nylon membranes are used by most off-the-shelf array providers such as Clontech Laboratories, Inc. (Palo Alto, CA), Genome Systems, Inc. (St. Louis, MO), and Research Genetics, Inc. (Huntsville, AL). Microarrays such as those produced by Affymetrix, Inc. (Santa Clara, CA), Incyte Pharmaceuticals, Inc. (Palo Alto, CA), and many do-it-yourself (DIY) arraying groups use glass wafers or slides. Although standard microscope slides may be used, they must be preprepared to facilitate sticking of the DNA to the glass. Several different

Address correspondence to J. Rockett, Reproductive Toxicology Division (MD-72), National Health and Environmental Effects Research Laboratory, U.S. EPA, Research Triangle Park, NC 27711 USA. Telephone: (919) 541-2678. Fax: (919) 541-4017. E-mail: rockett.john@epa.gov

The authors thank R. Kavlock for envisioning the application of array technology to toxicology at the U.S. Environmental Protection Agency. We also thank T. Wall and B. Deitz for administrative assistance.

This document has been reviewed in accordance with EPA policy and approved for publication. Mention of companies, trade names, or products does not signify endorsement of such by the EPA.

Received 23 March 1999; accepted 22 April 1999.

coatings have been successfully used, including silane and lysine. The coating of slides can easily be carried out in the laboratory, but many prefer the convenience of precoated slides available from suppliers.

Once the support matrix has been prepared, the DNA elements can be applied by several methods. Affymetrix, Inc., has developed a unique photolithographic technology for attaching oligonucleotides to glass wafers. More commonly, DNA is applied by either noncontact or contact printing. Noncontact printers can use thermal, solenoid, or piezoelectric technology to spray aliquots of solution onto the support matrix and may be used to produce slide or membrane-based arrays. Cartesian Technologies, Inc. (Irvine, CA) has developed nQUAD technology for use in its PixSys printers. The system couples a syringe pump with the microsolenoid valve, a combination that provides rapid quantitative dispensing of nanoliter volumes (down to 4.2 nL) over a variable volume range. A different approach to noncontact printing uses a solid pin and ring combination (Genetic MicroSystems, Inc., Woburn, MA). This system (Figure 1) allows a broader range of sample, including cell suspensions and particulates, because the printing head cannot be blocked up in the same way as a spray nozzle. Fluid transfer is controlled in this system primarily by the pin dimensions and the force of deposition, although the nature of the support matrix and the sample will also affect transfer to some degree.

In contact printing, the pin head is dipped in the sample and then touched to the support matrix to deposit a small aliquot. Split pins were one of the first contact-printing devices to be reported and are the suggested format for DIY arrayers, as described by Brown (*3*). Split pins are small metal pins with a precise groove cut vertically in the middle of the pin tip. In this system, 1–48 split pins are positioned in the pin-head. The split pins work by simple capillary action, not unlike a fountain pen—when the pin heads are dipped in the sample, liquid is drawn into the pin groove. A small (fixed) volume is then deposited each time the split pins are gently touched to the support matrix. Sample (100–500 pL depending on a variety of parameters) can be deposited on multiple slides before refilling is required, and array densities of > 2,500 spots/cm^2 may be produced. The deposit volume depends on the split size, sample fluidity, and the speed of printing. Split pins are relatively simple to produce and can be made in-house if a suitable machine shop is available. Alternatively, they can be obtained directly from companies such as TeleChem International, Inc. (Sunnyvale, CA).

Irrespective of their source, printers should be run through a preprint sequence prior to producing the actual experimental arrays: the first 100 or so spots of a new run tend to be somewhat variable. Factors effecting spot reproducibility include slide treatment homogeneity, sample differences, and instrument errors. Other factors that come into play include clean ejection of the drop and clogging (nQUAD printing) and mechanical variations and long-term alteration in print-head surface of solid and split pins. However, with careful preparation it is possible to get a coefficient of variance for spot reproducibility below 10%.

One potential printing problem is sample carryover. Repeated washing, blotting, and drying (vacuum) of print pins between samples is normally effective at reducing sample carryover to negligible amounts. Printing should also be carried out in a controlled environment. Humidified chambers are available in which to place printers. These help prevent dust contamination and produce a uniform drying rate, which is important in determining spot size, quality, and reproducibility.

In summary, although several printing technologies are available, none are particularly outstanding and the bottom line is that they are still in a relatively early stage of evolution.

Array Hybridization

The hybridization protocol is, practically speaking, relatively straightforward and those with previous experience in blotting should have little difficulty. Array hybridizations are, in essence, reverse Southern/Northern blots—instead of applying a labeled probe to the target population of DNA/RNA, the labeled population is applied to the probe(s). With membrane-based arrays, the control and treated mRNA populations are normally converted to cDNA and labeled with isotope (e.g., ^{33}P) in the process. These labeled populations are then hybridized independently to parallel or serial arrays and the hybridization signal is detected with a phosporimager. A less commonly used alternative to radioactive probes is enzymatic detection. The probe may be biotinylated, haptenylated, or have alkaline phosphatase/horseradish peroxidase attached. Hybridization is detected by enzymatic reaction yielding a color reaction (*4*). Differences in hybridization signals can be detected by eye or, more accurately, with the help of digital imaging and commercially available software. The labeling of the test populations for slide-based microarrays uses a slightly different approach. The probe typically consists of two samples of polyA$^+$ RNA (usually from a treated and a control population) that are converted to cDNA; in the process each is labeled with a different fluor. The independently labeled probes are then mixed together and hybridized to a single microarray slide and the resulting combined fluorescent signal is scanned. After normalization, it is possible to determine the ratio of fluorescent signals from a single hybridization of a slide-based microarray.

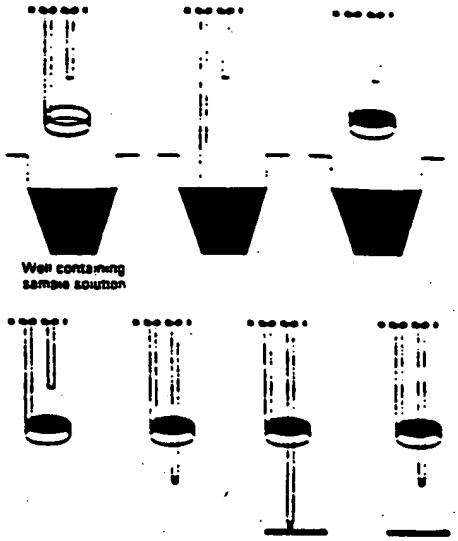


Figure 1. Genetic Microsystems (Woburn, MA) pin ring system for printing arrays. The pin ring combination consists of a circular open ring oriented parallel to the sample solution, with a vertical pin centered over the ring. When the ring is dipped into a solution and lifted, it withdraws an aliquot of sample held by surface tension. To spot the sample, the pin is driven down through the ring and a portion of the solution is transferred to the bottom of the pin. The pin continues to move downward until the pendant drop of solution makes contact with the underlying surface. The pin is then lifted, and gravity and surface tension cause deposition of the spot onto the array. Figure from Flowers et al. (*14*), with permission from Genetic Microsystems.

cDNA derived from control and treated populations of RNA is most commonly hybridized to arrays, although subtractive hybridization or differential display reactions may also be used. Fluorophore- or radiolabeled nucleotides are directly incorporated into the cDNA in the process of converting RNA to cDNA. Alternatively, 5′ end-labeled primers may be used for cDNA synthesis. These are labeled with a fluorophore for direct visualization of the hybridized array. Alternatively, biotin or a hapten may be attached to the primer, in which case fluor-labeled streptavidin or antibody must be applied before a signal can be generated. The most commonly used fluorophores at present are cyanine (Cy)3 and Cy5 (Amersham Pharmacia Biotech AB, Uppsala, Sweden). However, the relative expense of these fluorescent conjugates has driven a search for cheaper alternatives. Fluorescein, rhodamine, and Texas red have all been used, and companies such as Molecular Probes, Inc. (Eugene, OR) are developing a series of labeled nucleotides with a wide range of excitation and emission spectra which may prove to function as well as the Cy dyes.

Analysis of DNA Microarrays

Membrane-based arrays are normally analyzed on film or with a phosphorimager, whereas chip-based arrays require more specialized scanning devices. These can be divided into three main groups: the charge-coupled device camera systems, the nonconfocal laser scanners, and the confocal laser scanners. The advantages and disadvantages of each system are listed in Table 1.

Because a typical spot on a microarray can contain > 10^8 molecules, it is clear that a large variation in signal strength may occur. Current scanners cannot work across this many orders of magnitude (4 or 5 is more typical). However, the scanning parameters can normally be adjusted to collect more or less signal, such that two or three scans of the same array should permit the detection of rare and abundant genes.

When a microarray is scanned, the fluorescent images are captured by software normally included with the scanner. Several commercial suppliers provide additional software for quantifying array images, but the software tools are constantly evolving to meet the developing needs of researchers, and it is prudent to define one's own needs and clarify the exact capabilities of the software before its purchase. Issues that should be considered include the following:

- Can the software locate offset spots?
- Can it quantitate across irregular hybridization signals?
- Can the arrayed genes be programmed in for easy identification and location?
- Can the software connect via the Internet to databases containing further information on the gene(s) of interest?

One of the key issues raised at the workshop was the sensitivity of microarray technology. Experiments by General Scanning, Inc. (Watertown, MA), have shown that by using the Cy dyes and their scanner, signal can be detected down to levels of < 1 fluor molecule per square micrometer, which translates to detecting a rare message at approximately one copy per cell or less.

Array Applications

Although arrays are an emerging technology certain to undergo improvement and alteration, they have already been applied usefully to a number of model systems. Arrays are at their most powerful when they contain the entire genome of the species they are being used to study. For this reason, they have strong support among researchers utilizing yeast and *Caenorhabditis elegans* (5). The genomes of both of these species have been sequenced and, in the case of yeast, deposited onto arrays for examination of gene expression (6,7). With both of these species, it is relatively easy to perturb individual gene expression. Indeed, *C. elegans* knockouts can be made simply by soaking the worms in an antisense solution of the gene to be knocked out.

Table 1. Advantages and disadvantages of different microarray scanning systems.

	CCD camera system	Nonconfocal laser scanner	Confocal laser scanner
Advantages	Few moving parts	Relatively simple optics	Small depth of focus reduces artifacts
	Fast scanning of bright samples	—	May have high light collection efficiency
Disadvantages	Less appropriate for dim samples	Low light collection efficiency	Small depth of focus requires scanning precision
	Optical scatter can limit performance	Background artifacts not rejected	
		Resolution typically low	

CCD, charge-coupled device.
From Kawasaki (13).

By a process of systematic gene disruption, it is now possible to examine the cause and effect relationships between different genes in these simple organisms. This kind of approach should help elucidate biochemical pathways and genetic control processes, deconvolute polygenic interactions, and define the architecture of the cellular network. A simple case study of how this can be achieved was presented by Butow [University of Texas Southwestern Medical Center, Dallas, TX (Figure 2)]. Although it is the phenotypic result of a single gene knockout that is being examined, the effect of such perturbation will almost always be polygenic. Polygenic interactions will become increasingly important as researchers begin to move away from single gene systems when examining the nature of toxicologic responses to external stimuli. This is especially important in toxicology because the phenotype produced by a given environmental insult is never the result of the action of a single gene; rather, it is a complex interaction of one or multiple cellular pathways. Phenomena such as quantitative trait (the continuous variation of phenotype), epistasis (the effect of alleles of one or more genes on the expression of other genes), and penetrance (proportion of individuals of a given genotype that display a particular phenotype) will become increasingly evident and important as toxicologists push toward the ultimate goal of matching the responses of individuals to different environmental stimuli.

Analysis of the transcriptome (the expression level of all the genes in a given cell population) was a use of arrays addressed by several speakers. Unfortunately, current gene nomenclature is often confusing in that single genes are allocated multiple names (usually as a result of independent discovery by different laboratories), and there was a call for standardization of gene nomenclature. Nevertheless, once a transcriptome has been assembled it can then be transferred onto arrays and used to screen any chosen system. The EPA MicroArray Consortium (EPAMAC) is assembling testes transcriptomes for human, rat, and mouse. In a slightly different approach, Nuwaysir et al. (8) describes how the NIEHS assembled what is effectively a "toxicological transcriptome"—a library of human and mouse genes that have previously been proven or implicated in responses to toxicologic insults. Clontech Laboratories, Inc. (Palo Alto, CA), has begun a similar process by developing stress/toxicology filter arrays of rat, mouse, and human genes. Thus, rather than being tissue or cell specific, these stress/toxicology arrays can be used across a variety of model systems to look for alterations in the expression of toxicologically important genes and define the new field of toxicogenomics. The potential to identify toxicant families based on tissue- or cell-specific gene expression could revolutionize drug testing. These molecular signatures or fingerprints could not only point to the possible toxicity/carcinogenicity of newly discovered compounds (Figure 3), but also aid in elucidating their mechanism of action through identification of gene expression networks. By extension, such signatures could provide easily identifiable biomarkers to assess the degree, time, and nature of exposure.

DNA arrays are primarily a tool for examining differential gene expression in a given model. In this context they are referred to as closed systems because they lack the ability of other differential expression technologies, e.g., differential display and subtractive hybridization, to detect previously unknown genes not present on the array. This would appear to limit the power of DNA arrays to the imaginations and preconceptions of the researcher in selecting genes previously characterized and thought to be involved in the model system. However, the various genome sequencing projects have created a new category of sequence—the EST—that has partially mollified this deficiency. ESTs are cDNAs expressed in a given tissue that, although they may share some degree of sequence similarity to previously characterized genes, have not been assigned specific genetic identity. By incorporating EST clones into an array, it is possible to monitor the expression of these unknown genes. This can enable the identification of previously uncharacterized genes that may have biologic

significance in the model system. Filter arrays from Research Genetics and slide arrays from Incyte Pharmaceuticals both incorporate large numbers of ESTs from a variety of species.

A further use of microarrays is the identification of single nucleotide polymorphisms (SNPs). These genomic variations are abundant—they occur approximately every 1 kb or so—and are the basis of restriction fragment length polymorphism analysis used in forensic analysis. Affymetrix, Inc., designed chips that contain multiple repeats of the same gene sequence. Each position is present with all four possible bases. After the hybridization of the sample, the degree of hybridization to the different sequences can be measured and the exact sequence of the target gene deduced. SNPs are thought to be of vital importance in drug metabolism and toxicology. For example, single base differences in the regulatory region or active site of some genes can account for huge differences in the activity of that gene. Such SNPs are thought to explain why some people are able to metabolize certain xenobiotics better than others. Thus, arrays provide a further tool for the toxicologist investigating the nature of susceptible subpopulations and toxicologic response.

There are still many wrinkles to be ironed out before arrays become a standard tool for toxicologists. The main issues raised at the workshop by those with hands-on experience were the following:

- Expense: the cost of purchasing/contracting this technology is still too great for many individual laboratories.
- Clones: the logistics of identifying, obtaining, and maintaining a set of nonredundant, noncontaminated, sequence-verified, species/cell/tissue/field-specific clones.
- Use of inbred strains: where whole-organism models are being used, the use of inbred strains is important to reduce the potentially confusing effects of the individual variation typically seen in outbred populations.
- Probe: the need for relatively large amounts of RNA, which limits the type of sample (e.g., biopsy) that can be used. Also, different RNA extraction methods can give different results.
- Specificity: the ability to discriminate accurately between closely related genes (e.g., the cytochrome p450 family) and splice variants.
- Quantitation: the quantitation of gene expression using gene arrays is still open to debate. One reason for this is the different incorporation of the labeling dyes. However, the main difficulty lies in knowing what to normalize against. One option is to include a large number of so-called housekeeping genes in the array. However, the expression of these genes often change depending on the tissue and the toxicant, so it is necessary to characterize the expression of these genes in the model system before utilizing them. This is clearly not a viable option when screening multiple new compounds. A second option is to include on the array genes from a nonrelated species (e.g., a plant gene on an animal array) and to spike the probe with synthetic RNA(s) complementary to the gene(s).
- Reproducibility: this is sometimes questionable, and a figure of approximately two or three repeats was used as the minimum number required to confirm initial findings. Again, however, most people advocated the use of Northern blots or reverse transcriptase PCR to confirm findings.
- Sensitivity: concerns were voiced about the number of target molecules that must be present in a sample for them to be detected on the array.
- Efficiency: reproducible identification of 1.5- to 2-fold differences in expression was reported, although the number of genes that undergo this level of change and remain undetected is open to debate. It is important that this level of detection be ultimately achieved because it is commonly perceived that some important transcription factors and their regulators respond at such low levels. In most cases, 3- to 5-fold was the minimum change that most were happy to accept.
- Bioinformatics: perhaps the greatest concern was how to accurately interpret the data with the greatest accuracy and efficiency. The biggest headache is trying to identify networks of gene expression that are common to different treatments or doses. The amount of data from a single experiment is huge. It may be that, in the future, several groups individually equipped with specialized software algorithms for studying their favorite genes or gene systems will be able to share the same hybridized chips. Thus, arrays could usher in a new perspective on collaboration and the sharing of data.

EPAMAC

Perhaps the main reason most scientists are unable to use array technology is the high cost involved, whether buying off-the-shelf membranes, using contract printing services, or

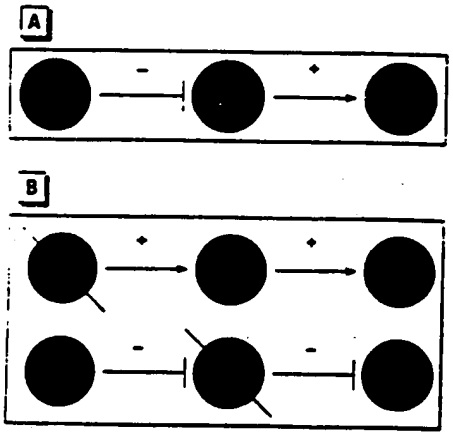


Figure 2. Potential effects of gene knockout within positively and negatively regulated gene expression networks. i_1 is limiting in wild type for expression of i_2. (*A*) A simple, two-component, linear regulatory network operating on gene i_2 where i_1 is a positive effector of i_2 and j_n is either a positive or negative effector of i_1. This network could be deduced by examining the consequence of (*B*) deleting j_n on the expression of i_1 and i_2 where the expression of i_2 would be decreased or increased depending on whether j_n was a positive or negative regulator. These and other connected components of even greater complexity could be revealed by genome-wide expression analysis. From Butow (*15*).

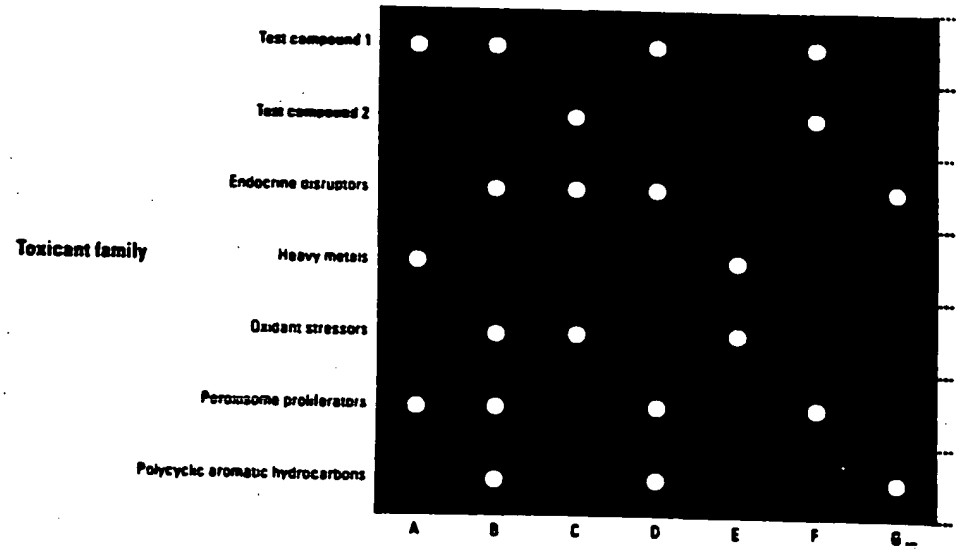


Figure 3. Gene expression profiles—also called fingerprints or signatures—of known toxicants or toxicant families may, in the future, be used to identify the potential toxicity of new drugs, etc. In this example, the genetic signature of test compound 1 is identical to that of known peroxisome proliferators, whereas that of test compound 2 does not match any known toxicant family. Based on these results, test compound 2 would be retained for further testing and test compound 1 would be eliminated.

producing chips in-house. In view of this, researchers at the RTD/NHEERL initiated the EPAMAC. This consortium brings together scientists from the EPA and a number of extramural labs with the aim of developing microarray capability through the sharing of resources and data. EPAMAC researchers are primarily interested in the developmental and toxicologic changes seen in testicular and breast tissue, and a portion of the workshop was set aside for EPAMAC members to share their ideas on how the experimental application of microarrays could facilitate their research. One of the central areas of interest to EPAMAC members is the effect of xenobiotics on male fertility and reproductive health. Of greatest concern is the effect of exposure during critical periods of development and germ cell differentiation (*9*), and how this may compromise sperm counts and quality following sexual maturation (*10*). As well as spermatogenic tissue, there is also interest in how residual mRNA found in mature sperm (*11*) could be used as an indicator of previous xenobiotic effects (it is easier to obtain a semen sample than a testicular biopsy). Arrays will be used to examine and compare the effect of exposure to heat and chemicals in testicular and epididymal gene expression profiles, with the aim of establishing relationships/associations between changes in developmental landmarks and the effects on sperm count and quality. Cluster, pattern, and other analysis of such data should help identify hidden relationships between genes that may reveal potential mechanisms of action and uncover roles for genes with unknown functions.

Summary

The full impact of DNA arrays may not be seen for several years, but the interest shown at this regional workshop indicates the high level of interest that they foster. Apart from educating and advertising the various technologies in this field, this workshop brought together a number of researchers from the Research Triangle Park area who are already using DNA arrays. The interest in sharing ideas and experiences led to the initiation of a Triangle array user's group.

Array technology is still in its infancy. This means that the hardware is still improving and there is no current consensus for standard procedures, quantitation, and interpretation. Consistency in spotting and scanning arrays is not yet optimized, and this is one of the most critical requirements of any experiment. In addition, one of the dark regions of array technology—strife in the courts over who owns what portions of it—has further muddled the future and is a potential barrier toward the development of consensus procedures.

Perhaps the greatest hurdle for the application of arrays is the actual interpretation of data. No specialists in bioinformatics attended the workshop, largely because they are rare and because as yet no one seems clear on the best method of approaching data analysis and interpretation. Cross-referencing results from multiple experiments (time, dose, repeats, different animals, different species) to identify commonly expressed genes is a great challenge. In most cases, we are still a long way from understanding how the expression of gene *X* is related to the expression of gene *Y*, and ordering gene expression to delineate causal relationships.

To the ordinary scientist in the typical laboratory, however, the most immediate problem is a lack of affordable instrumentation. One can purchase premade membranes at relatively affordable prices. Although these may be useful in identifying individual genes to pursue in more detail using other methods, the numbers that would be required for even a small routine toxicology experiment prohibit this as a truly viable approach. For the toxicologist, there is a need to carry out multiple experiments—dose responses, time curves, multiple animals, and repeats. Glass-based DNA arrays are most attractive in this context because they can be prepared in large batches from the same DNA source and accommodate control and treated samples on the same chip. Another problem with current off-the-shelf arrays is that they often do not contain one or more of the particular genes a group is interested in. One alternative is to obtain and/or produce a set of custom clones and have contract printing of membranes or slides carried out by a company such as Genomic Solutions, Inc. (Ann Arbor, MI). This approach is less expensive than having out capital for one's own entire system, although at some point it might make economic sense to print one's own arrays.

Finally, DNA arrays are currently a team effort. They are a technology that uses a wide range of skills including engineering, statistics, molecular biology, chemistry, and bioinformatics. Because most individuals are skilled in only one or perhaps two of these areas, it appears that success with arrays may be best expected by teams of collaborators consisting of individuals having each of these skills.

Those considering array applications may be amused or goaded on by the following quote from *Fortune* magazine (*12*):

> Microprocessors have reshaped our economy, spawned vast fortunes and changed the way we live. Gene chips could be even bigger.

Although this comment may have been designed to excite the imagination rather than accurately reflect the truth, it is fair to say that the age of functional genomics is upon us. DNA arrays look set to be an important tool in this new age of biotechnology and will likely contribute answers to some of toxicology's most fundamental questions.

Speakers

Cindy Afshari, NIEHS
Linda Birnbaum, U.S. EPA
Ron Butow, University of Texas Southwestern Medical Center
Alex Chenchik, Clontech Laboratories, Inc.
David Dix, U.S. EPA
Abdel Elkahloun, Research Genetics, Inc.
Sue Fenton, U.S. EPA
Norman Hecht, University of Pennsylvania
Pat Hurban, Paradigm Genetics, Inc.
Bob Kavlock, U.S. EPA
Erme Kawasaki, General Scanning, Inc.
Steve Krawetz, Wayne State University
Nick Mace, Genetic Microsystems, Inc.
Scott Mordecai, Affymetrix, Inc.
Kevin Morgan, Glaxo Wellcome, Inc.
Elaine Poplin, Research Genetics, Inc.
Don Rose, Cartesian Technologies, Inc.
Jim Samet, U.S. EPA
Sam Ward, University of Arizona
Jeff Welch, U.S. EPA
Reen Wu, University of California at Davis
Tim Zacharewski, Michigan State University

References and Notes

1. The chipping forecast. Nat Genet 21(Suppl 1):3–60 (1999).
2. National Center for Biotechnology Information. The Unigene System. Available: www.ncbi.nlm.nih.gov/Schuler/UniGene [cited 22 March 1999].
3. Brown PO. The Brown Lab. Available: http://cmgm.Stanford.edu/pbrown [cited 22 March 1999].
4. Chen JJ, Wu R, Yang PC, Huang JY, Sher YP, Han MH, Kao WC, Lee PJ, Chiu TF, Chang F, et al. Profiling expression patterns and isolating differentially expressed genes by cDNA microarray system with colorimetry detection. Genomics 51:313–324 (1998).
5. Ward S. DNA Microarray Technology to Identify Genes Controlling Spermatogenesis. Available: www.mcb.arizona.edu/wardlab/microarray.html [cited 22 March 1999].
6. Marton MJ, DeRisi JL, Bennett HA, Iyer VR, Meyer MR, Roberts CJ, Stoughton R, Burchard J, Slade D, Dai H, et al. Drug target validation and identification of secondary drug target effects using DNA microarrays. Nat Med 4:1293–1301 (1998).
7. Brown PO. The Full Yeast Genome on a Chip. Available: http://cmgm.stanford.edu/pbrown/yeastchip.html [cited 22 March 1999].
8. Nuwaysir EF, Bittner M, Trent J, Barrett JC, Afshari CA. Microarrays and toxicology: the advent of toxicogenomics. Mol Carcinog 24(3):153–159 (1999).
9. Hecht NB. Molecular mechanisms of male germ cell differentiation. Bioessays 20:555–561 (1998).
10. Zacharewski TR. Timothy R. Zacharewski. Available: www.bch.msu.edu/faculty/zacher.htm [cited 22 March 1999].
11. Kramer JA, Krawetz SA. RNA in spermatozoa: implications for the alternative haploid genome. Mol Hum Reprod 3:473–478 (1997).
12. Stipp D. Gene chip breakthrough. Fortune, March 31:56–73 (1997).
13. Kawasaki E (General Scanning Instruments, Inc., Watertown, MA). Unpublished data.
14. Flowers P, Overbeck J, Mace ML Jr, Pagliughi FM, Eggers WJE, Yonkers H, Honkanen P, Montagu J, Rose SD. Development and Performance of a Novel Microarraying System Based on Surface Tension Forces. Available: http://www.geneticmicro.com/resources/html/coldspring.html [cited 22 March 1999].
15. Butow R (University of Texas Medical Center, Dallas, TX). Unpublished data.

XENOBIOTICA, 1999, VOL. 29, NO. 7, 655–691

Differential gene expression in drug metabolism and toxicology: practicalities, problems and potential

JOHN C. ROCKETT†, DAVID J. ESDAILE‡ and G. GORDON GIBSON*

Molecular Toxicology Laboratory, School of Biological Sciences, University of Surrey, Guildford, Surrey, GU2 5XH, UK

Received January 8, 1999

1. An important feature of the work of many molecular biologists is identifying which genes are switched on and off in a cell under different environmental conditions or subsequent to xenobiotic challenge. Such information has many uses, including the deciphering of molecular pathways and facilitating the development of new experimental and diagnostic procedures. However, the student of gene hunting should be forgiven for perhaps becoming confused by the mountain of information available as there appears to be almost as many methods of discovering differentially expressed genes as there are research groups using the technique.

2. The aim of this review was to clarify the main methods of differential gene expression analysis and the mechanistic principles underlying them. Also included is a discussion on some of the practical aspects of using this technique. Emphasis is placed on the so-called 'open' systems, which require no prior knowledge of the genes contained within the study model. Whilst these will eventually be replaced by 'closed' systems in the study of human, mouse and other commonly studied laboratory animals, they will remain a powerful tool for those examining less fashionable models.

3. The use of suppression-PCR subtractive hybridization is exemplified in the identification of up- and down-regulated genes in rat liver following exposure to phenobarbital, a well-known inducer of the drug metabolizing enzymes.

4. Differential gene display provides a coherent platform for building libraries and microchip arrays of 'gene fingerprints' characteristic of known enzyme inducers and xenobiotic toxicants, which may be interrogated subsequently for the identification and characterization of xenobiotics of unknown biological properties.

Introduction

It is now apparent that the development of almost all cancers and many non-neoplastic diseases are accompanied by altered gene expression in the affected cells compared to their normal state (Hunter 1991, Wynford-Thomas 1991, Vogelstein and Kinzler 1993, Semenza 1994, Cassidy 1995, Kleinjan and Van Hegningen 1998). Such changes also occur in response to external stimuli such as pathogenic micro-organisms (Rohn *et al.* 1996, Singh *et al.* 1997, Griffin and Krishna 1998, Lunney 1998) and xenobiotics (Sewall *et al.* 1995, Dogra *et al.* 1998, Ramana and Kohli 1998), as well as during the development of undifferentiated cells (Hecht 1998, Rudin and Thompson 1998, Schneider-Maunoury *et al.* 1998). The potential medical and therapeutic benefits of understanding the molecular changes which occur in any given cell in progressing from the normal to the 'altered' state are enormous. Such profiling essentially provides a 'fingerprint' of each step of a

* Author for correspondence; e-mail: g.gibson@surrey.ac.uk
† Current Address: US Environmental Protection Agency, National Health and Environmental Effects, Research Laboratory, Reproductive Toxicology Division, Research Triangle Park, NC 27711, USA.
‡ Rhone-Poulenc Agrochemicals, Toxicology Department, Sophia-Antipolis, Nice, France.

Xenobiotica ISSN 0049–8254 print/ISSN 1366–5928 online © 1999 Taylor & Francis Ltd
http://www.tandf.co.uk/JNLS/xen.htm http://www.taylorandfrancis.com/JNLS/xen.htm

cell's development or response and should help in the elucidation of specific and sensitive biomarkers representing, for example, different types of cancer or previous exposure to certain classes of chemicals that are enzyme inducers.

In drug metabolism, many of the xenobiotic-metabolizing enzymes (including the well-characterized isoforms of cytochrome P450) are inducible by drugs and chemicals in man (Pelkonen *et al.* 1998), predominantly involving transcriptional activation of not only the cognate cytochrome P450 genes, but additional cellular proteins which may be crucial to the phenomenon of induction. Accordingly, the development of methodology to identify and assess the full complement of genes that are either up- or down-regulated by inducers are crucial in the development of knowledge to understand the precise molecular mechanisms of enzyme induction and how this relates to drug action. Similarly, in the field of chemical-induced toxicity, it is now becoming increasingly obvious that most adverse reactions to drugs and chemicals are the result of multiple gene regulation, some of which are causal and some of which are casually-related to the toxicological phenomenon *per se*. This observation has led to an upsurge in interest in gene-profiling technologies which differentiate between the control and toxin-treated gene pools in target tissues and is, therefore, of value in rationalizing the molecular mechanisms of xenobiotic-induced toxicity. Knowledge of toxin-dependent gene regulation in target tissues is not solely an academic pursuit as much interest has been generated in the pharmaceutical industry to harness this technology in the early identification of toxic drug candidates, thereby shortening the developmental process and contributing substantially to the safety assessment of new drugs. For example, if the gene profile in response to say a testicular toxin that has been well-characterized *in vivo* could be determined in the testis, then this profile would be representative of all new drug candidates which act via this specific molecular mechanism of toxicity, thereby providing a useful and coherent approach to the early detection of such toxicants. Whereas it would be informative to know the identity and functionality of all genes up/down regulated by such toxicants, this would appear a longer term goal, as the majority of human genes have not yet been sequenced, far less their functionality determined. However, the current use of gene profiling yields a *pattern* of gene changes for a xenobiotic of unknown toxicity which may be matched to that of well-characterized toxins, thus alerting the toxicologist to possible *in vivo* similarities between the unknown and the standard, thereby providing a platform for more extensive toxicological examination. Such approaches are beginning to gain momentum, in that several biotechnology companies are commercially producing 'gene chips' or 'gene arrays' that may be interrogated for toxicity assessment of xenobiotics. These chips consist of hundreds/thousands of genes, some of which are degenerate in the sense that not all of the genes are mechanistically-related to any one toxicological phenomenon. Whereas these chips are useful in broad-spectrum screening, they are maturing at a substantial rate, in that gene arrays are now becoming more specific, e.g. chips for the identification of changes in growth factor families that contribute to the aetiology and development of chemically-induced neoplasias.

Although documenting and explaining these genetic changes presents a formidable obstacle to understanding the different mechanisms of development and disease progression, the technology is now available to begin attempting this difficult challenge. Indeed, several 'differential expression analysis' methods have been developed which facilitate the identification of gene products that demonstrate

altered expression in cells of one population compared to another. These methods have been used to identify differential gene expression in many situations, including invading pathogenic microbes (Zhao *et al.* 1998), in cells responding to extracellular and intracellular microbial invasion (Duguid and Dinauer 1990, Ragno *et al.* 1997, Maldarelli *et al.* 1998), in chemically treated cells (Syed *et al.* 1997, Rockett *et al.* 1999), neoplastic cells (Liang *et al.* 1992, Chang and Terzaghi-Howe 1998), activated cells (Gurskaya *et al.* 1996, Wan *et al.* 1996), differentiated cells (Hara *et al.* 1991, Guimaraes *et al.* 1995a, b), and different cell types (Davis *et al.* 1984, Hedrick *et al.* 1984, Xhu *et al.* 1998). Although differential expression analysis technologies are applicable to a broad range of models, perhaps their most important advantage is that, in most cases, absolutely no prior knowledge of the specific genes which are up- or down-regulated is required.

The field of differential expression analysis is a large and complex one, with many techniques available to the potential user. These can be categorized into several methodological approaches, including:

(1) Differential screening,
(2) Subtractive hybridization (SH) (includes methods such as chemical cross-linking subtraction—CCLS, suppression-PCR subtractive hybridization—SSH, and representational difference analysis—RDA),
(3) Differential display (DD),
(4) Restriction endonuclease facilitated analysis (including serial analysis of gene expression—SAGE—and gene expression fingerprinting—GEF),
(5) Gene expression arrays, and
(6) Expressed sequence tag (EST) analysis.

The above approaches have been used successfully to isolate differentially expressed genes in different model systems. However, each method has its own subtle (and sometimes not so subtle) characteristics which incur various advantages and disadvantages. Accordingly, it is the purpose of this review to clarify the mechanistic principles underlying the main differential expression methods and to highlight some of the broader considerations and implications of this very powerful and increasingly popular technique. Specifically, we will concentrate on the so-called 'open' systems, namely those which do not require any knowledge of gene sequences and, therefore, are useful for isolating unknown genes. Two 'closed' systems (those utilising previously identified gene sequences), EST analysis and the use of DNA arrays, will also be considered briefly for completeness. Whilst emphasis will often be placed on suppression PCR subtractive hybridization (SSH, the approach employed in this laboratory), it is the aim of the authors to highlight, wherever possible, those areas of common interest to those who use, or intend to use, differential gene expression analysis.

Differential cDNA library screening (DS)

Despite the development of multiple technological advances which have recently brought the field of gene expression profiling to the forefront of molecular analysis, recognition of the importance of differential gene expression and characterization of differentially expressed genes has existed for many years. One of the original approaches used to identify such genes was described 20 years ago by St John and Davis (1979). These authors developed a method, termed 'differential plaque filter

hybridization', which was used to isolate galactose-inducible DNA sequences from yeast. The theory is simple: a genomic DNA library is prepared from normal, unstimulated cells of the test organism/tissue and multiple filter replicas are prepared. These replica blots are probed with radioactively (or otherwise) labelled complex cDNA probes prepared from the control and test cell mRNA populations. Those mRNAs which are differentially expressed in the treated cell population will show a positive signal only on the filter probed with cDNA from the treated cells. Furthermore, labelled cDNA from different test conditions can be used to probe multiple blots, thereby enabling the identification of mRNAs which are only up-regulated under certain conditions. For example, St John and Davis (1979) screened replica filters with acetate-, glucose- and galactose-derived probes in order to obtain genes induced specifically by galactose metabolism. Although groundbreaking in its time this method is now considered insensitive and time-consuming, as up to 2 months are required to complete the identification of genes which are differentially expressed in the test population. In addition, there is no convenient way to check that the procedure has worked until the whole process has been completed.

Subtractive Hybridization (SH)

The developing concept of differential gene expression and the success of early approaches such as that described by St John and Davis (1979) soon gave rise to a search for more convenient methods of analysis. One of the first to be developed was SH, numerous variations of which have since been reported (see below). In general, this approach involves hybridization of mRNA/cDNA from one population (tester) to excess mRNA/cDNA from another (driver), followed by separation of the unhybridized tester fraction (differentially expressed) from the hybridized common sequences. This step has been achieved physically, chemically and through the use of selective polymerase chain reaction (PCR) techniques.

Physical separation

Original subtractive hybridization technology involved the physical separation of hybridized common species from unique single stranded species. Several methods of achieving this have been described, including hydroxyapatite chromatography (Sargent and Dawid 1983), avidin-biotin technology (Duguid and Dinauer 1990) and oligodT-latex separation (Hara *et al.* 1991). In the first approach, common mRNA species are removed by cDNA (from test cells)-mRNA (from control cells) subtractive hybridization followed by hydroxyapatite chromatography, as hydroxyapatite specifically adsorbs the cDNA-mRNA hybrids. The unabsorbed cDNA is then used either for the construction of a cDNA library of differentially expressed genes (Sargent and Dawid 1983, Schneider *et al.* 1988) or directly as a probe to screen a preselected library (Zimmerman *et al.* 1980, Davis *et al.* 1984, Hedrick *et al.* 1984). A schematic diagram of the procedure is shown in figure 1.

Less rigorous physical separation procedures coupled with sensitivity enhancing PCR steps were later developed as a means to overcome some of the problems encountered with the hydroxyapatite procedure. For example, Daguid and Dinauer (1990) described a method of subtraction utilizing biotin-affinity systems as a means to remove hybridized common sequences. In this process, both the control and tester mRNA populations are first converted to cDNA and an adaptor ('oligovector',

)NA sequences from
:pared from normal,
le filter replicas are
·r otherwise) labelled
mRNA populations.
d cell population will
·om the treated cells.
:an be used to probe
; which are only up-
)avis (1979) screened
bes in order to obtain
groundbreaking in its
nsuming, as up to 2
ıich are differentially
/enient way to check
:n completed.

d the success of early
9) soon gave rise to a
t to be developed was
:e below). In general,
ıe population (tester)
›y separation of the
hybridized common
and through the use

e physical separation
:ies. Several methods
tite chromatography
i and Dinauer 1990)
: approach, common
A (from control cells)
ography, as hydroxy-
ınabsorbed cDNA is
fferentially expressed
lirectly as a probe to
l. 1984, Hedrick *et al.*
:re 1.
sensitivity enhancing
›me of the problems
Daguid and Dinauer
ty systems as a means
›oth the control and
daptor ('oligovector',

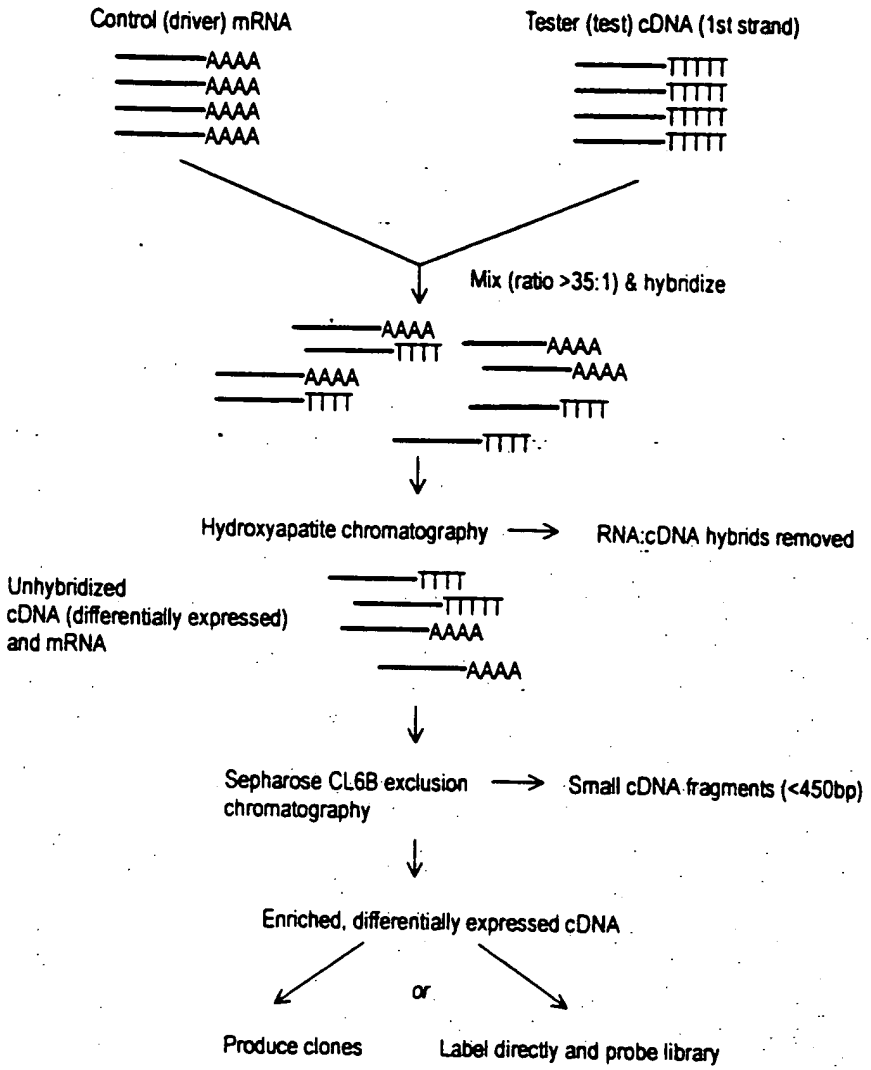


Figure 1. The hydroxyapatite method of subtractive hybridization. cDNA derived from the treated/altered (tester) population is mixed with a large excess of mRNA from the control (driver) population. Following hybridization, mRNA-cDNA hybrids are removed by hydroxyapatite chromatography. The only cDNAs which remain are those which are differentially expressed in the treated/altered population. In order to facilitate the recovery of full length clones, small cDNA fragments are removed by exclusion chromatography. The remaining cDNAs are then cloned into a vector for sequencing, or labelled and used directly to probe a library, as described by Sargent and Dawid (1983).

containing a restriction site) ligated to both sides. Both populations are then amplified by PCR, but the driver cDNA population is subsequently digested with the adaptor-containing restriction endonuclease. This serves to cleave the oligovector and reduce the amplification potential of the control population. The digested control population is then biotinylated and an excess mixed with tester cDNA. Following denaturation and hybridization, the mix is applied to a biocytin column (streptavidin may also be used) to remove the control population, including heteroduplexes formed by annealing of common sequences from the tester population. The procedure is repeated several times following the addition of fresh

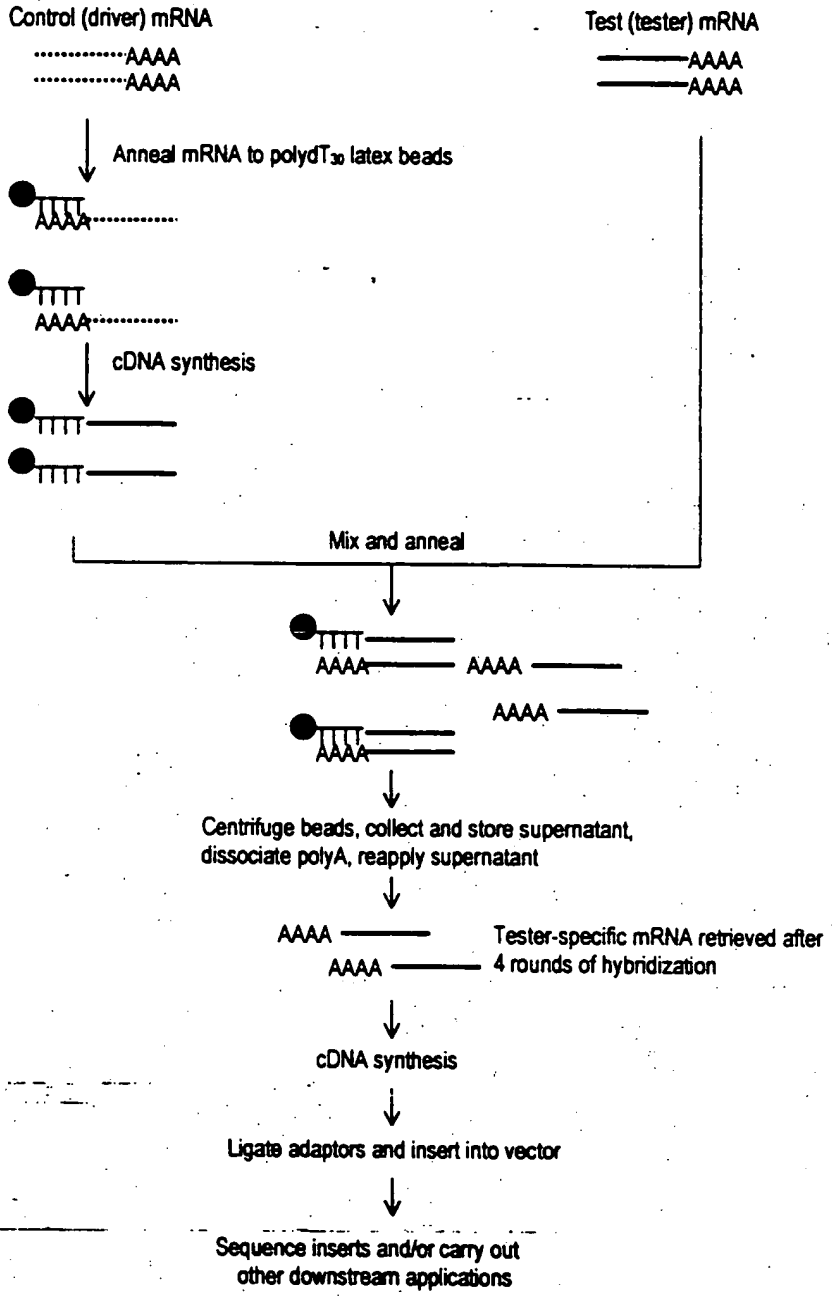


Figure 2. The use of oligodT_{30} latex to perform subtractive hybridization. mRNA extracted from the control (driver) population is converted to anchored cDNA using polydT oligonucleotides attached to latex beads. mRNA from the treated/altered (tester) population is repeatedly hybridized against an excess of the anchored driver cDNA. The final population of mRNA is tester specific and can be converted into cDNA for cloning and other downstream applications, as described by Hara *et al.* (1991).

control cDNA. In order to further enrich those species differentially expressed in the tester cDNA, the subtracted tester population is amplified by PCR following every second subtraction cycle. After six cycles of subtraction (three reamplification steps) the reaction mix is ligated into a vector for further analysis.

In a slightly different approach, Hara *et al.* (1991) utilized a method whereby oligo(dT_{30}) primers attached to a latex substrate are used to first capture mRNA extracted from the control population. Following 1st strand cDNA synthesis, the RNA strand of the heteroduplexes is removed by heat denaturation and centrifugation (the cDNA-oligotex-dT_{30} forms a pellet and the supernatant is removed). A quantity of tester mRNA is then repeatedly hybridized to the immobilized control (driver) cDNA (which is present in 20-fold excess). After several rounds of hybridization the only mRNA molecules left in the tester mRNA population are those which are not found in the driver cDNA-oligotex-dT_{30} population. These tester-specific mRNA species are then converted to cDNA and, following the addition of adaptor sequences, amplified by PCR. The PCR products are then ligated into a vector for further analysis using restriction sites incorporated into the PCR primers. A schematic illustration of this subtraction process is shown in figure 2.

However, all these methods utilising physical separation have been described as inefficient due to the requirement for large starting amounts of mRNA, significant loss of material during the separation process and a need for several rounds of hybridization. Hence, new methods of differential expression analysis have recently been designed to eliminate these problems.

Chemical Cross-Linking Subtraction (CCLS)

In this technique, originally described by Hampson *et al.* (1992), driver mRNA is mixed with tester cDNA (1st strand only) in a ratio of > 20:1. The common sequences form cDNA:mRNA hybrids, leaving the tester specific species as single stranded cDNA. Instead of physically separating these hybrids, they are inactivated chemically using 2,5 diaziridinyl-1,4-benzoquinone (DZQ). Labelled probes are then synthesized from the remaining single stranded cDNA species (unreacted mRNA species remaining from the driver are not converted into probe material due to specificity of Sequenase T7 DNA polymerase used to make the probe) and used to screen a cDNA library made from the tester cell population. A schematic diagram of the system is shown in figure 3.

It has been shown that the differentially expressed sequences can be enriched at least 300-fold with one round of subtraction (Hampson *et al.* 1992), and that the technique should allow isolation of cDNAs derived from transcripts that are present at less than 50 copies per cell. This equates to genes at the low end of intermediate abundance (see table 1). The main advantages of the CCLS approach are that it is rapid, technically simple and also produces fewer false positives than other differential expression analysis methods. However, like the physical separation protocols, a major drawback with CCLS is the large amount of starting material required (at least 10 μg RNA). Consequently, the technique has recently been refined so that a renewable source of RNA can be generated. The degenerate random oligonucleotide primed (DROP) adaptation (Hampson *et al.* 1996, Hampson and Hampson 1997) uses random hexanucleotide sequences to prime solid phase-synthesized cDNA. Since each primer includes a T7 polymerase promotor sequence

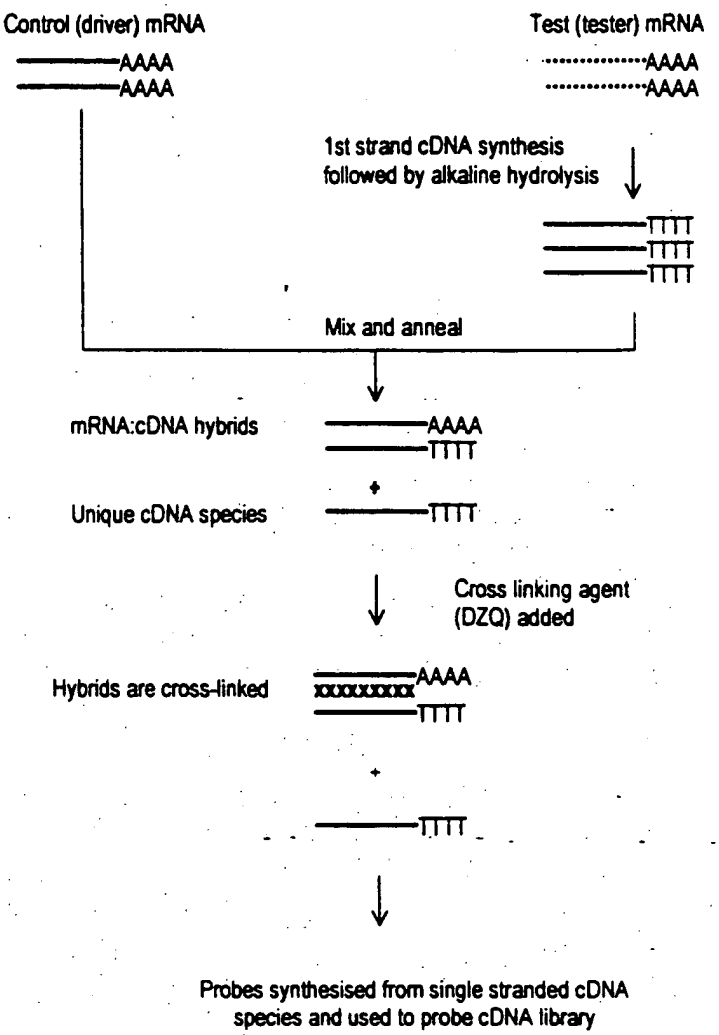


Figure 3. Chemical cross-linking subtraction. Excess driver mRNA is mixed with 1st strand tester cDNA. The common sequences form mRNA:cDNA hybrids which are cross linked with 2,5 diaziridinyl-1,4-benzoquinone (DZQ) and the remaining cDNA sequences are differentially expressed in the tester population. Probes are made from these sequences using Sequenase 2.0 DNA polymerase, which lacks reverse transcriptase activity and, therefore, does not react with the remaining mRNA molecules from the driver. The labelled probes are then used to screen a cDNA library for clones of differentially expressed sequences. Adapted from Walter *et al.* (1996), with permission.

Table 1. The abundance of mRNA species and classes in a typical mammalian cell.

mRNA class	Copies of each species/cell	No. of mRNA species in class	Mean % of each species in class	Mean mass (ng) of each species/μg total RNA
Abundant	12000	4	3.3	1.65
Intermediate	300	500	0.08	0.04
Rare	15	11000	0.004	0.002

Modified from Bertioli *et al.* (1995).

at the 5′ end, the final pool of random cDNA fragments is a PCR-renewable cDNA population which is representative of the expressed gene pool and can be used to synthesize sense RNA for use as driver material. Furthermore, if the final pool of random cDNA fragments is reamplified using biotinylated T7 primer and random hexamer, the product can be captured with streptavidin beads and the antisense strand eluted for use as tester. Since both target and driver can be generated from the same DROP product, subtraction can be performed in both directions (i.e. for up- and down-regulated species) between two different DROP products.

Representational Difference Analysis (RDA)

RDA of cDNA (Hubank and Schatz 1994) is an extension of the technique originally applied to genomic DNA as a means of identifying differences between two complex genomes (Lisitsyn *et al.* 1993). It is a process of subtraction and amplification involving subtractive hybridization of the tester in the presence of excess driver. Sequences in the tester that have homologues in the driver are rendered unamplifiable, whereas those genes expressed only in the tester retain the ability to be amplified by PCR. The procedure is shown schematically in figure 4.

In essence, the driver and tester mRNA populations are first converted to cDNA and amplified by PCR following the ligation of an adaptor. The adaptors are then removed from both populations and a new (different) adaptor ligated to the amplified tester population only. Driver and tester populations are next melted and hybridized together in a ratio of 100:1. Following hybridization, only tester:tester homohybrids have 5′ adaptors at each end of the DNA duplex and can, thus, be filled in at both 3′ ends. Hence, only these molecules are amplified exponentially during the subsequent PCR step. Although tester:driver heterohybrids are present, they only amplify in a linear fashion, since the strand derived from the driver has no adaptor to which the primer can bind. Driver:driver heterohybrids have no adaptors and, therefore, are not amplified. Single stranded molecules are digested with mung bean nuclease before a further PCR-enrichment of the tester:tester homohybrids. The adaptors on the amplified tester population are then replaced and the whole process repeated a further two or three times using an increasing excess of driver (Hubank and Shatz used a tester:driver ratio of 1:400, 1:80000 and 1:800000 for the second, third and fourth hybridizations, respectively). Different adaptors are ligated to the tester between successive rounds of hybridization and amplification to prevent the accumulation of PCR products that might interfere with subsequent amplifications. The final display is a series of differentially expressed gene products easily observable on an ethidium bromide gel.

The main advantages of RDA are that it offers a reproducible and sensitive approach to the analysis of differentially expressed genes. Hubank and Schatz (1994) reported that they were able to isolate genes that were differentially expressed in substantially less than 1 % of the cells from which the tester is derived. Perhaps the main drawback is that multiple rounds of ligation, hybridization, amplifiation and digestion are required. The procedure is, therefore, lengthier than many other differential display approaches and provides more opportunity for operator-induced error to occur. Although the generation of false positives has been noted, this has been solved to some degree by O'Neill and Sinclair (1997) through the use of HPLC-purified adaptors. These are free of the truncated adaptors which appear to be a major source of the false positive bands. A very similar technique to RDA, termed linker capture subtraction (LCS) was described by Yang and Sytowski (1996).

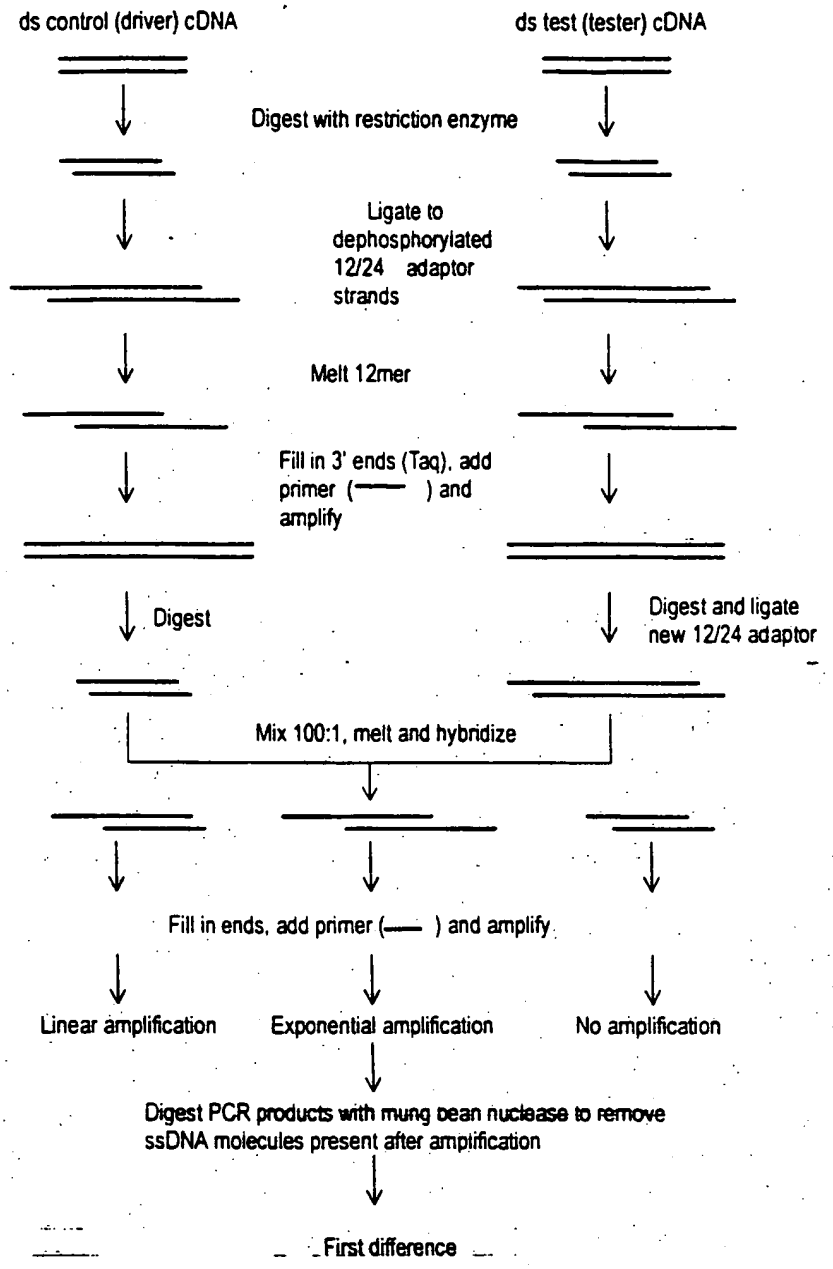


Figure 4. The representational difference analysis (RDA) technique. Driver and tester cDNA are digested with a 4-cutter restriction enzyme such as *Dpn*II. The 1st set of 12/24 adaptor strands (oligonucleotides) are ligated to each other and the digested cDNA products. The 12mer is subsequently melted away and the 3'ends filled in using Taq DNA polymerase. Each cDNA population is then amplified using PCR, following which the 1st set of adaptors is removed with *Dpn*II. A second set of 12/24 adaptor strands is then added to the amplified tester cDNA population, after which the tester is hybridized against a large excess of driver. The 12mer adaptors are melted and the 3' ends filled in as before. PCR is carried out with primers identical to the new 24mer adaptor. Thus, the only hybridization products which are exponentially amplified are those which are tester:tester combinations. Following PCR, ssDNA products are removed with mung bean nuclease, leaving the 'first difference product'. This is digested and a third set of 12/24 adaptors added before repeating the subtraction process from the hybridization stage. The process is repeated to the 3rd or 4th difference product, as described by Lisitsyn *et al.* (1993) and Hubank and Schatz (1994).

r) cDNA

Digest and ligate
new 12/24 adaptor

ification

e

iver and tester cDNA are
et of 12/24 adaptor strands
A products. The 12mer is
\ polymerase. Each cDNA
of adaptors is removed with
he amplified tester cDNA
cess of driver. The 12mer
l out with primers identical
s which are exponentially
PCR, ssDNA products are
uct'. This is digested and a
cess from the hybridization
described by Lisitsyn *et al.*

Suppression PCR Subtractive Hybridization (SSH)

The most recent adaptation of the SH approach to differential expression analysis was first described by Diatchenko *et al.* (1996) and Gurskaya *et al.* (1996). They reported that a 1000–5000 fold enrichment of rare cDNAs (equivalent to isolating mRNAs present at only a few copies per cell) can be obtained without the need for multiple hybridizations/subtractions. Instead of physical or chemical removal of the common sequences, a PCR-based suppression system is used (see figure 5).

In SSH, excess driver cDNA is added to two portions of the tester cDNA which have been ligated with different adaptors. A first round of hybridization serves to enrich differentially expressed genes and equalize rare and abundant messages. Equalization occurs since reannealing is more rapid for abundant molecules than for rarer molecules due to the second order kinetics of hybridization (James and Higgins 1985). The two primary hybridization mixes are then mixed together in the presence of excess driver and allowed to hybridize further. This step permits the annealing of single stranded complementary sequences which did not hybridize in the primary hybridization, and in doing so generates templates for PCR amplification. Although there are several possible combinations of the single stranded molecules present in the secondary hybridization mix, only one particular combination (differentially expressed in the tester cDNA composed of complimentary strands having different adaptors) can amplify exponentially.

Having obtained the final differential display, two options are available if cloning of cDNAs is desired. One is to transform the whole of the final PCR reaction into competent cells. Transformed colonies can then be isolated and their inserts characterized by sequencing, restriction analysis or PCR. Alternatively, the final PCR products can be resolved on a gel and the individual bands excised, reamplified and cloned. The first approach is technically simpler and less time consuming. However, ligation/transformation reactions are known to be biased towards the cloning of smaller molecules, and so the final population of clones will probably not contain a representative selection of the larger products. In addition, although equalization theoretically occurs, observations in this laboratory suggest that this is by no means perfectly accomplished. Consequently, some gene species are present in a higher number than others and this will be represented in the final population of clones. Thus, in order to obtain a substantial proportion of those gene species that actually demonstrate differential expression in the tester population, the number of clones that will have to be screened after this step may be substantial. The second approach is initially more time consuming and technically demanding. However, it would appear to offer better prospects for cloning larger and low abundance gel products. In addition, one can incorporate a screening step that differentiates different products of different sequences but of the same size (HA-staining, see later). In this way, a good idea of the final number of clones to be isolated and identified can be achieved.

An alternative (or even complementary) approach is to use the final differential display reaction to screen a cDNA library to isolate full length clones for further characterization, or a DNA array (see later) to quickly identify known genes. SSH has been used in this laboratory to begin characterization of the short-term gene expression profiles of enzyme-inducers such as phenobarbital (Rockett *et al.* 1997) and Wy-14,643 (Rockett *et al.* unpublished observations). The isolation of differentially expressed genes in this manner enables the construction of a fingerprint

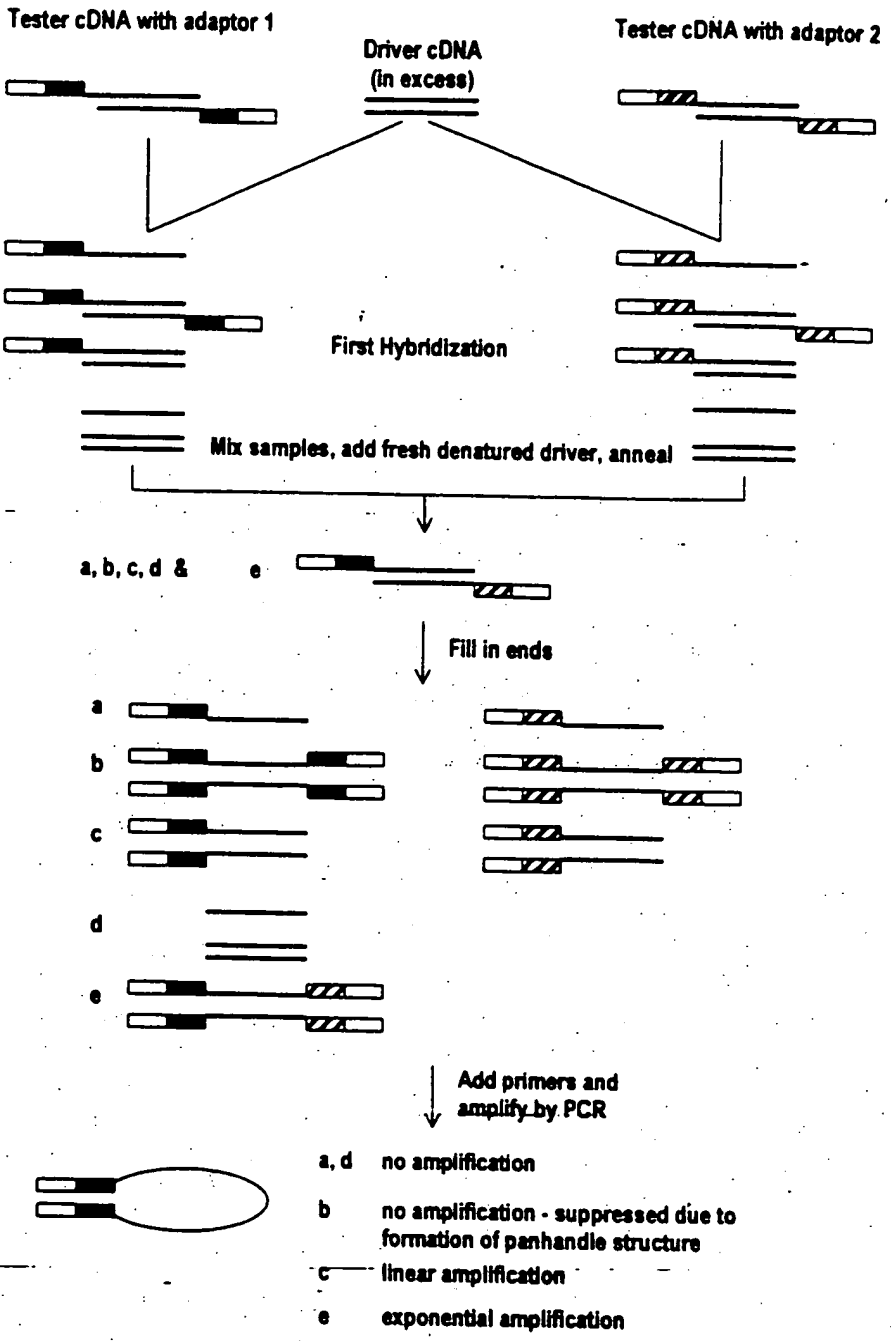


Figure 5. PCR-select cDNA subtraction. In the primary hybridization, an excess of driver cDNA is added to each tester cDNA population. The samples are heat denatured and allowed to hybridize for between 3 and 8 h. This serves two purposes: (1) to equalize rare and abundant molecules; and (2) to enrich for differentially expressed sequences—cDNAs that are not differentially expressed form type c molecules with the driver. In the secondary hybridization, the two primary hybridizations are mixed together without denaturing. Fresh denatured driver can also be added at this point to allow further enrichment of differentially expressed sequences. Type e molecules are formed in this secondary hybridization which are subsequently amplified using two rounds of PCR. The final products can be visualized on an agarose gel, labelled directly or cloned into a vector for downstream manipulation. As described by Diatchenko *et al.* (1996) and Gurskaya *et al.* (1996), with permission.

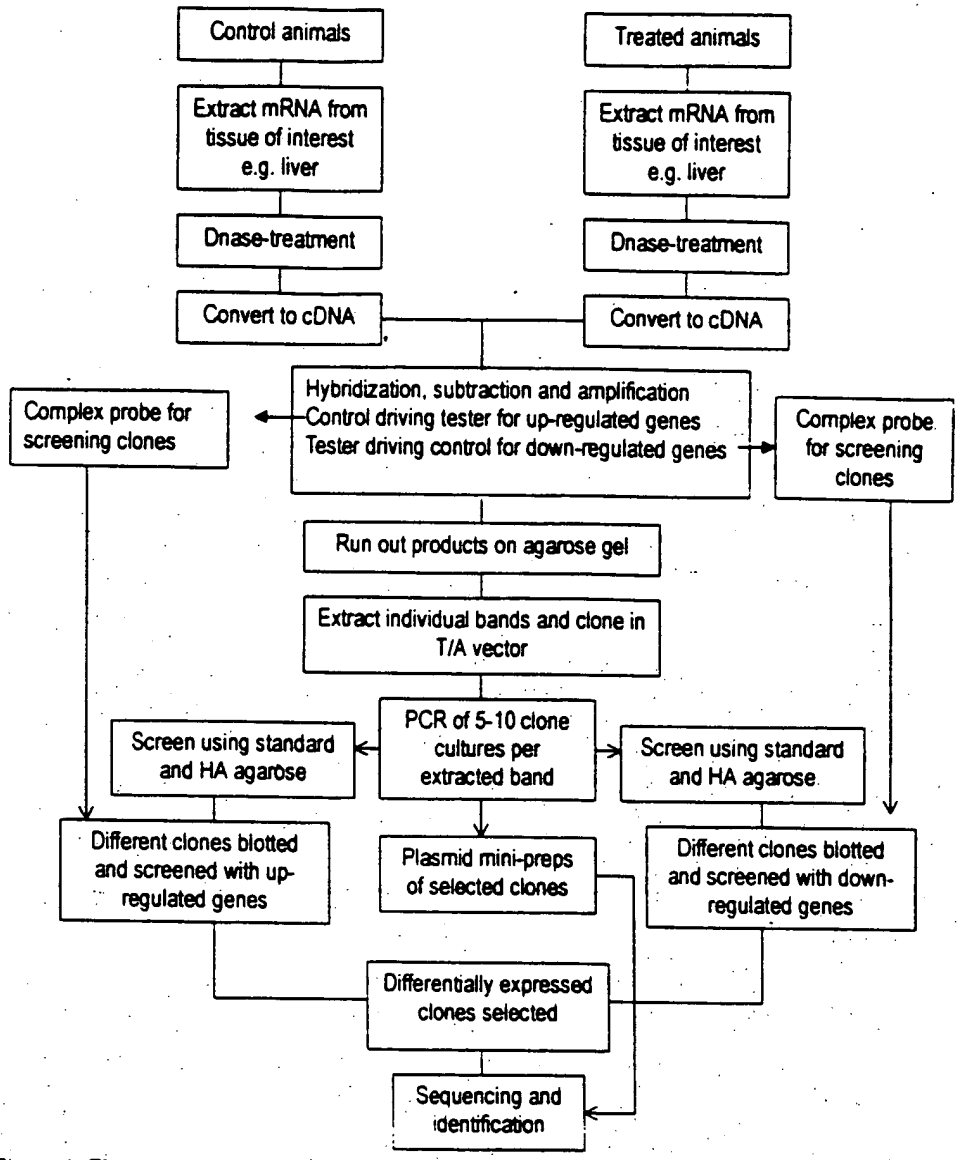


Figure 6. Flow diagram showing method used in this laboratory to isolate and identify clones of genes which are differentially expressed in rat liver following short term exposure to the enzyme inducers, phenobarbital and Wy-14,643.

of expressed genes which are unique to each compound and time/dose point. Such information could be useful in short-term characterization of the toxic potential of new compounds by comparing the gene-expression profiles they elicit with those produced by known inducers. Figure 6 shows a flow diagram of the method used to isolate, verify and clone differentially expressed genes, and figure 7 shows expression profiles obtained from a typical SSH experiment. Subsequent sub-cloning of the individual bands, sequencing and gene data base interrogation reveals many genes which are either up- or down-regulated by phenobarbital in the rat (tables 2 and 3).

One of the advantages in using the SSH approach is that no prior knowledge is required of which specific genes are up/down-regulated subsequent to xenobiotic

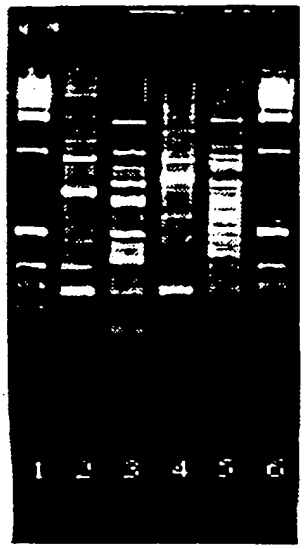


Figure 7. SSH display patterns obtained from rat liver following 3-day treatment with WY-14,643 or phenobarbital. mRNA extracted from control and treated livers was used to generate the differential displays using the PCR-Select cDNA subtraction kit (Clontech). Lane: 1—1kb ladder; 2—genes upregulated following Wy,14–643 treatment; 3—genes downregulated following Wy,14–643 treatment; 4—genes upregulated following phenobarbital treatment; 5—genes downregulated following phenobarbital treatment; 6—1kb ladder. Reproduced from Rockett *et al.* (1997), with permission.

exposure, and an almost complete complement of genes are obtained. For example, the peroxisome proliferator and non-genotoxic hepatocarcinogen Wy,14,643, up-regulates at least 28 genes and down-regulates at least 15 in the rat (a sensitive species) and produces 48 up- and 37 down-regulated genes in the guinea pig, a resistant species (Rockett, Swales, Esda and Gibson, unpublished observations). One of these genes, CD81, was up-regulated in the rat and down-regulated in the guinea pig following Wy-14,643 treatment. CD81 (alternatively named TAPA-1) is a widely expressed cell surface protein which is involved in a large number of cellular processes including adhesion, activation, proliferation and differentiation (Levy *et al.* 1998). Since all of these functions are altered to some extent in the phenomena of hepatomegaly and non-genotoxic hepatocarcinogenesis, it is intriguing, and probably mechanistically-relevant, that CD81 expression is differentially regulated in a resistant and susceptible species. However, the down-side of this approach is that the majority of genes can be sequenced and matched to database sequences, but the latter are predominantly expressed sequence tags or genes of completely unknown function, thus partially obscuring a realistic overall assessment of the critical genes of genuine biological interest. Notwithstanding the lack of complete funtional identification of altered gene expression, such gene profiling studies essentially provides a 'molecular fingerprint' in response to xenobiotic challenge, thereby serving as a mechanistically-relevant platform for further detailed investigations.

Differential Display (DD)

Originally described as 'RNA fingerprinting by arbitrarily primed PCR' (Liang and Pardee 1992) this method is now more commonly referred to as 'differential

Table 2. Genes up-regulated in rat liver following 3-day exposure to phenobarbital.

Band number (approximate size in bp)	Highest sequence similarity	FASTA-EMBL gene identification
5 (1300)	93.5%	CYP2B1
7 (1000)	95.1%	Preproalbumin Serum albumin mRNA
8 (950)	98.3%	NCI-CGAP-Pr1 *H. sapiens* (EST)
10 (850)	95.7%	CYP2B1
11 (800)	Clone 1 94.9%	CYP2B1
	Clone 2 75.3%	CYP2B2
12 (750)	93.8%	TRPM-2 mRNA Sulfated glycoprotein
15 (600)	92.9%	Preproalbumin Serum albumin mRNA
16 (55)	Clone 1 95.2%	CYP2B1
	Clone 2 93.6%	Haptoglobulin mRNA partial alpha
21 (350)	99.3%	18S, 5.8S & 28S rRNa

Bands 1–4, 6, 9, 13, 14, and 17–20 are shown to be false positives by dot blot anaylsis and, therefore, are not sequenced. Derived from Rockett *et al.* (1997). It should be noted that the above genes do not represent the complete spectrum of genes which are up-regulated in rat liver by phenobarbital, but simply represents the genes sequenced and identified to date.

Table 3. Genes down-regulated in rat liver following 3-day exposure to phenobarbital.

Band number (approximate size in bp)	Highest sequence similarity	FASTA-EMBL gene identification
1 (1500)	95.3%	3-oxoacyl-CoA thiolase
2 (1200)	92.3%	Hemopoxin mRNA
3 (1000)	91.7%	Alpha-2u-globulin mRNA
7 (700)	Clone 1 77.2%	*M.musculus* Cl inhibitor
	Clone 2 94.5%	Electron transfer flavoprotein
	Clone 3 91.0%	*M. musculus* Topoisomerase 1 (Topo 1)
8 (650)	Clone 1 86.9%	Soares 2NbMT *M. musculus* (EST)
	Clone 2 96.2%	Alpha-2u-globulin (s-type) mRNA
9 (600)	Clone 1 86.9%	Soares mouse NML *M. musculus* (EST)
	Clone 2 82.0%	Soares p3NMF 19.5 *M. musculus* (EST)
10 (550)	73.8%	Soares mouse NML *M. musculus* (EST)
11 (525)	95.7%	NCI-CGAP-Pr1 *H. sapiens* (EST)
12 (375)	100.0%	Ribosomal protein
13 (23)	Clone 1 97.2%	Soares mouse embryo NbME135 (EST)
	Clone 2 100.0%	Fibrinogen B-beta-chain
	Clone 3 100.0%	Apolipoprotein E gene
14 (170)	96.0%	Soares p3NMF19.5 *M. musculus* (EST)
15 (140)	97.3%	Stratagene mouse testis (EST)
Others: (300)	96.7%	*R. norvegicus* RASP 1 mRNA
(275)	93.1%	Soares mouse mammary gland (EST)

EST = Expressed sequence tag. Bands 4–6 were shown to be false positives by dot blot analysis and, therefore, were not sequenced. Derived from Rockett *et al.* (1997). It should be noted that the above genes do not represent the complete spectrum of genes which are down-regulated in rat liver by phenobarbital, but simply represents the genes sequenced and identified to date.

display' (DD). In this method, all the mRNA species in the control and treated cell populations are amplified in separate reactions using reverse transcriptase-PCR (RT-PCR). The products are then run side-by-side on sequencing gels. Those bands which are present in one display only, or which are much more intense in one

display compared to the other, are differentially expressed and may be recovered for further characterization. One advantage of this system is the speed with which it can be carried out—2 days to obtain a display and as little as a week to make and identify clones.

Two commonly used variations are based on different methods of priming the reverse transcription step (figure 8). One is to use an oligo dT with a 2-base 'anchor' at the 3'-end, e.g. 5′ (dT_{11})CA 3′ (Liang and Pardee 1992). Alternatively, an arbitrary primer may be used for 1st strand cDNA synthesis (Welsh *et al.* 1992). This variant of RNA fingerprinting has also been called 'RAP' (RNA Arbitrarily Primed)-PCR. One advantage of this second approach is that PCR products may be derived from anywhere in the RNA, including open reading frames. In addition, it can be used for mRNAs that are not polyadenylated, such as many bacterial mRNAs (Wong and McClelland 1994). In both cases, following reverse transcription and denaturation, second strand cDNA synthesis is carried out with an arbitrary primer (*arbitrary* primers have a single base at each position, as compared to *random* primers, which contain a mixture of all four bases at each position). The resulting PCR, thus, produces a series of products which, depending on the system (primer length and composition, polymerase and gel system), usually includes 50–100 products per primer set (Band and Sager 1989). When a combination of different dT-anchors and arbitrary primers are used, almost all mRNA species from a cell can be amplified. When the cDNA products from two different populations are analysed side by side on a polyacrylamide gel, differences in expression can be identified and the appropriate bands recovered for cloning and further analysis.

Although DD is perhaps the most popular approach used today for identifying differentially expressed genes, it does suffer from several perceived disadvantages:

(1) It may have a strong bias towards high copy number mRNAs (Bertioli *et al.* 1995), although this has been disputed (Wan *et al.* 1996) and the isolation of very low abundance genes may be achieved in certain circumstances (Guimeraes *et al.* 1995a).
(2) The cDNAs obtained often only represent the extreme 3′ end of the mRNA (often the 3′-untranslated region), although this may not always be the case (Guimeraes *et al.* 1995a). Since the 3′ end is often not included in Genbank and shows variation between organisms. cDNAs identified by DD cannot always be matched with their genes, even if they have been identified.
(3) The pattern of differential expression seen on the display often cannot be reproduced on Northern blots, with false positives arising in up to 70% of cases (Sun *et al.* 1994). Some adaptations have been shown to reduce false positives, including the use of two reverse transcriptases (Sung and Denman 1997), comparison of uninduced and induced cells over a time course (Burn *et al.* 1994) and comparison of DDPCR-products from two uninduced and two induced lines (Sompayrac *et al.* 1995). The latter authors also reported that the use of cytoplasmic RNA rather then total RNA reduces false positives arising from nuclear RNA that is not transported to the cytoplasm.

Further details of the background, strengths and weaknesses of the DD technique can be obtained from a review by McClelland *et al.* (1996) and from articles by Liang *et al.* (1995) and Wan *et al.* (1996).

d may be recovered for
peed with which it can
k to make and identify

ethods of priming the
with a 2-base 'anchor'
92). Alternatively, an
is (Welsh *et al.* 1992).
AP' (RNA Arbitrarily
PCR products may be
frames. In addition, it
ıany bacterial mRNAs
erse transcription and
ith an arbitrary primer
compared to *random*
ɔsition). The resulting
on the system (primer
ıally includes 50–100
ɪbination of different
species from a cell can
ɔpulations are analysed
ı can be identified and
lysis.
l today for identifying
ceived disadvantages:

ıRNAs (Bertioli *et al.*
ıd the isolation of very
stances (Guimeraes *et*

3′ end of the mRNA
ot always be the case
iuded in Genbank and
DD cannot always be
ed.
play often cannot be
in up to 70 % of cases
reduce false positives,
and Denman 1997),
urse (Burn *et al.* 1994)
ıced and two induced
ported that the use of
positives arising from

aknesses of the DD
t al. (1996) and from

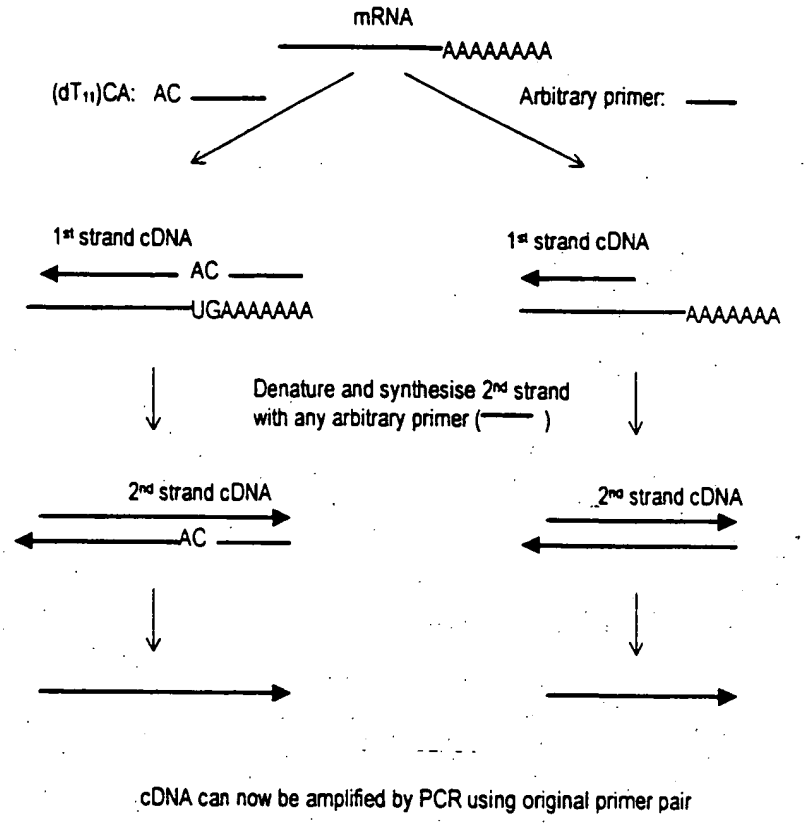


Figure 8. Two approaches to differential display (DD) analysis. 1st strand synthesis can be carried out either with a polydT_{11}NN primer (where N = G, C or A) or with an arbitrary primer. The use of different combinations of G, C and A to anchor the first strand polydT primer enables the priming of the majority of polyadenylated mRNAs. Arbitrary primers may hybridize at none, one or more places along the length of the mRNA, allowing 1st strand cDNA synthesis to occur at none, one or more points in the same gene. In both cases, 2nd strand synthesis is carried out with an arbitrary primer. Since these arbitrary primers for the 2nd strand may also hybridize to the 1st strand cDNA in a number of different places, several different 2nd strand products may be obtained from one binding point of the 1st strand primer. Following 2nd strand synthesis, the original set of primers is used to amplify the second strand products, with the result that numerous gene sequences are amplified.

Restriction endonuclease-facilitated analysis of gene expression

Serial Analysis of Gene Expression (SAGE)

A more recent development in the field of differential display is SAGE analysis (Velculescu *et al.* 1995). This method uses a different approach to those discussed so far and is based on two principles. Firstly, in more than 95 % of cases, short nucleotide sequences ('tags') of only nine or 10 base pairs provide sufficient information to identify their gene of origin. Secondly, concatonation (linking together in a series) of these tags allows sequencing of multiple cDNAs within a single clone. Figure 9 shows a schematic representation of the SAGE process. In this procedure, double stranded cDNA from the test cells is synthesized with a biotinylated polydT primer. Following digestion with a commonly cutting (4bp recognition sequence) restriction enzyme ('anchoring enzyme'), the 3′ ends of the cDNA population are captured with streptavidin beads. The captured population is

split into two and different adaptors ligated to the 5′ ends of each group. Incorporated into the adaptors is a recognition sequence for a type IIS restriction enzyme—one which cuts DNA at a defined distance (< 20 bp) from its recognition sequence. Hence, following digestion of each captured cDNA population with the IIS enzyme, the adaptors plus a short piece of the captured cDNA are released. The two populations are then ligated and the products amplified. The amplified products are cleaved with the original anchoring enzyme, religated (concatomers are formed in the process) and cloned. The advantage of this system is that hundreds of gene tags can be identified by sequencing only a few clones. Furthermore, the number of times a given transcript is identified is a quantitative measurement of that gene's abundance in the original population, a feature which facilitates identification of differentially expressed genes in different cell populations.

Some disadvantages of SAGE analysis include the technical difficulty of the method, a large amount of accurate sequencing is required, biased towards abundant mRNAs, has not been validated in the pharmaco/toxicogenomic setting and has only been used to examine well known tissue differences to date.

Gene Expression Fingerprinting (GEF)

A different capture/restriction digest approach for isolating differentially expressed genes has been described by Ivanova and Belyavsky (1995). In this method, RNA is converted to cDNA using biotinylated oligo(dT) primers. The cDNA population is then digested with a specific endonuclease and captured with magnetic streptavidin microbeads to facilitate removal of the unwanted 5′ digestion products. The use of restricted 3′-ends alone serves to reduce the complexity of the cDNA fragment pool and helps to ensure that each RNA species is represented by not more than one restriction product. An adaptor is ligated to facilitate subsequent amplification of the captured population. PCR is carried out with one adaptor-specific and one biotinylated polydT primer. The reamplified population is recaptured and the non-biotinylated strands removed by alkaline dissociation. The non-biotinylated strand is then resynthesized using a different adaptor-specific primer in the presence of a radiolabelled dNTP. The labelled immobilized 3′ cDNA ends are next sequentially treated with a series of different restriction endonucleases and the products from each digestion analysed by PAGE. The result is a fingerprint composed of a number of ladders (equal to the number of sequential digests used). By comparing test versus control fingerprints, it is possible to identify differentially expressed products which can then be isolated from the gel and cloned. The advantages of this procedure are that it is very robust and reproducible, and the authors estimate that 80–93% of cDNA molecules are involved in the final fingerprint. The disadvantage is that polyacrylamide gels can rarely resolve more than 300–400 bands, which compares poorly to the 1000 or more which are estimated to be produced in an average experiment. The use of 2-D gels such as those described by Uitterlinden *et al.* (1989) and Hatada *et al.* (1991) may help to overcome this problem.

A similar method for displaying restriction endonuclease fragments was later described by Prashar and Weissman (1996). However, instead of sequential digestion of the immobolized 3′-terminal cDNA fragments, these authors simply compared the profiles of the control and treated populations without further manipulation.

:h group. Incorporated
striction enzyme—one
recognition sequence.
n with the IIS enzyme,
re released. The two
amplified products are
atomers are formed in
hundreds of gene tags
e, the number of times
ement of that gene's
itates identification of

inical difficulty of the
ased towards abundant
nomic setting and has
date.

solating differentially
avsky (1995). In this
igo(dT) primers. The
ase and captured with
unwanted 5′ digestion
the complexity of the
ecies is represented by
o facilitate subsequent
ut with one adaptor-
plified population is
line dissociation. The
erent adaptor-specific
immobilized 3′ cDNA
triction endonucleases
result is a fingerprint
quential digests used).
identify differentially
gel and cloned. The
reproducible, and the
involved in the final
n rarely resolve more
) or more which are
e of 2-D gels such as
al. (1991) may help to

e fragments was later
instead of sequential
these authors simply
ions without further

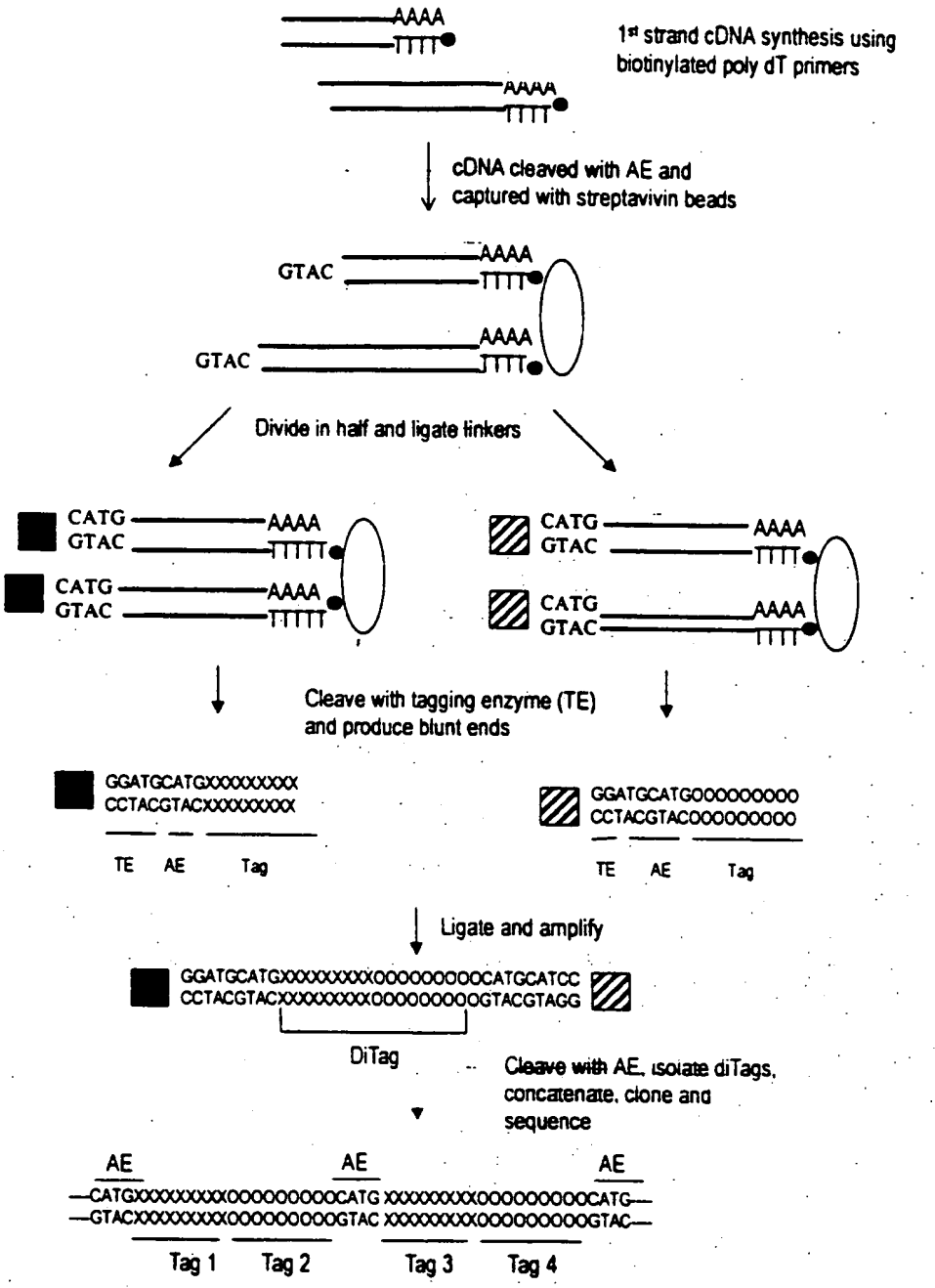


Figure 9. Serial analysis of gene expression (SAGE) analysis. cDNA is cleaved with an anchoring enzyme (AE) and the 3′ ends captured using streptavidin beads. The cDNA pool is divided in half and each portion ligated to a different linker, each containing a type IIS restriction site (tagging enzyme, TE). Restriction with the type IIS enzyme releases the linker plus a short length of cDNA (XXXXX and OOOOO indicate nucleotides of different tags). The two pools of tags are then ligated and amplified using linker-specific primers. Following PCR, the products are cleaved with the AE and the ditags isolated from the linkers using PAGE. The ditags are then ligated (during which process, concatenization occurs) and cloned into a vector of choice for sequencing. After Velculescu *et al.* (1995), with permission.

DNA arrays

'Open' differential display systems are cumbersome in that it takes a great deal of time to extract and identify candidate genes and then confirm that they are indeed up- or down-regulated in the treated compared to the control tissue. Normally, the latter process is carried out using Northern blotting or RT-PCR. Even so, each of the aforementioned steps produce a bottleneck to the ultimate goal of rapid analysis of gene expression. These problems will likely be addressed by the development of so-called DNA arrays (e.g. Gress *et al.* 1992, Zhao *et al.* 1995, Schena *et al.* 1996), the introduction of which has signalled the next era in differential gene expression analysis. DNA arrays consist of a gridded membrane or glass 'chips' containing hundreds or thousands of DNA spots, each consisting of multiple copies of part of a known gene. The genes are often selected based on previously proven involvement in oncogenesis, cell cycling, DNA repair, development and other cellular processes. They are usually chosen to be as specific as possible for each gene and animal species. Human and mouse arrays are already commercially available and a few companies will construct a personalized array to order, for example Clontech Laboratories and Research Genetics Inc. The technique is rapid in that hundreds or even thousands of genes can be spotted on a single array, and that mRNA/cDNA from the test populations can be labelled and used directly as probe. When analysed with appropriate hardware and software, arrays offer a rapid and quantitative means to assess differences in gene expression between two cell populations. Of course, there can only be identification and quantitation of those genes which are in the array (hence the term 'closed' system). Therefore, one approach to elucidating the molecular mechanisms involved in a particular disease/development system may be to combine an open and closed system—a DNA array to directly identify and quantitate the expression of known genes in mRNA populations, and an open system such as SSH to isolate unknown genes which are differentially expressed.

One of the main advantages of DNA arrays is the huge number of gene fragments which can be put on a membrane—some companies have reported gridding up to 60000 spots on a single glass 'chip' (microscope slide). These high density chip-based micro-arrays will probably become available as mass-produced off-the-shelf items in the near future. This should facilitate the more rapid determination of differential expression in time and dose-response experiments. Aside from their high cost and the technical complexities involved in producing and probing DNA arrays, the main problem which remains, especially with the newer micro-array (gene-chip) technologies, is that results are often not wholly reproducible between arrays. However, this problem is being addressed and should be resolved within the next few years.

EST databases as a means to identify differentially expressed genes

Expressed sequence tags (ESTs) are partial sequences of clones obtained from cDNA libraries. Even though most ESTs have no formal identity (putative identification is the best to be hoped for), they have proven to be a rapid and efficient means of discovering new genes and can be used to generate profiles of gene-expression in specific cells. Since they were first described by Adams *et al.* (1991), there has been a huge explosion in EST production and it is estimated that there are now well over a million such sequences in the public domain, representing over half

of all human genes (Hillier *et al.* 1996). This large number of freely available sequences (both sequence information and clones are normally available royalty-free from the originators) has enabled the development of a new approach towards differential gene expression analysis as described by Vasmatzis *et al.* (1998). The approach is simple in theory: EST databases are first searched for genes that have a number of related EST sequences from the target tissue of choice, but none or few from non-target tissue libraries. Programmes to assist in the assembly of such sets of overlapping data may be developed in-house or obtained privately or from the internet. For example, the Institute for Genomic Research (TIGR, found at http://www.tigr.org) provides many software tools free of charge to the scientific community. Included amongst these is the TIGR assembler (Sutton *et al.* 1995), a tool for the assembly of large sets of overlapping data such as ESTs, bacterial artificial chromosomes (BAC)s, or small genomes. Candidate EST clones representing different genes are then analysed using RNA blot methods for size and tissue specificity and, if required, used as probes to isolate and identify the full length cDNA clone for further characterization. In practice however, the method is rather more involved, requiring bioinformatic and computer analysis coupled with confirmatory molecular studies. Vasmatzis *et al.* (1998) have described several problems in this fledgling approach, such as separating highly homologous sequences derived from different genes and an overemphasis of specificity for some EST sequences. However, since these problems will largely be addressed by the development of more suitable computer algorithms and an increased completeness of the EST database, it is likely that this approach to identifying differentially expressed genes may enjoy more patronage in the future.

Problems and potential of differential expression techniques

The holistic or single cell approach?

When working with *in vivo* models of differential expression, one of the first issues to consider must be the presence of multiple cell types in any given specimen. For example, a liver sample is likely to contain not only hepatocytes, but also (potentially) Ito cells, bile ductule cells, endothelial cells, various immune cells (e.g. lymphocytes, macrophages and Kupffer cells) and fibroblasts. Other tissues will each have their own distinctive cell populations. Also, in the case of neoplastic tissue, there are almost always normal, hyperplastic and/or dysplastic cells present in a sample. One must, therefore, be aware that genes obtained from a differential display experiment performed on an animal tissue model may not necessarily arise exclusively from the intended 'target' cells, e.g. hepatocytes/neoplastic cells. If appropriate, further analyses using immunohistochemistry, *in situ* hybridization or *in situ* RT-PCR should be used to confirm which cell types are expressing the gene(s) of interest. This problem is probably most acute for those studying the differential expression of genes in the development of different cell types, where there is a need to examine homologous cell populations. The problem is now being addressed at the National Cancer Institute (Bethesda, MD, USA) where new microdisection techniques have been employed to assist in their gene analysis programme, the Cancer Genome Anatomy Project (CGAP) (For more information see web site: http://www.ncbi.nlm.nih.gov/ncicgap/intro.html). There are also separation techniques available that utilise cell-specific antigens as a means to isolate target cells,

e.g. fluorescence activated cell sorting (FACS) (Dunbar *et al.* 1998, Kas-Deelen *et al.* 1998) and magnetic bead technology (Richard *et al.* 1998, Rogler *et al.* 1998).

However, those taking a holistic approach may consider this issue unimportant. There is an equally appropriate view that all those genes showing altered expression within a compromized tissue should be taken into consideration. After all, since all tissues are complex mixes of different, interacting cell types which intimately regulate each other's growth and development, it is clear that each cell type could in some way contribute (positively or negatively) towards the molecular mechanisms which lie behind responses to external stimuli or neoplastic growth. It is perhaps then more informative to carry out differential display experiments using *in vivo* as opposed to *in vitro* models, where uniform populations of identical cells probably represent a partial, skewed or even inaccurate picture of the molecular changes that occur.

The incidence and possible implications of inter-individual biological variation should be considered in any approach where whole animal models are being used. It is clear that individuals (humans and animals) respond in different ways to identical stimuli. One of the best characterized examples is the debrisoquine oxidation polymorphism, which is mediated by cytochrome CYP2D6 and determines the pharmacokinetics of many commonly prescribed drugs (Lennard 1993, Meyer and Zanger 1997). The reasons for such differences are varied and complex, but allelic variations, regulatory region polymorphisms and even physical and mental health can all contribute to observed differences in individual responses. Careful thought should, therefore, be given to the specific objectives of the study and to the possible value of pooling starting material (tissue/mRNA). The effect of this can be beneficial through the ironing out of exaggerated responses and unimportant minor fluctuations of (mechanistically) irrelevant genes in individual animals, thus providing a clearer overall picture of the general molecular mechanisms of the response. However, at the same time such minor variations may be of utmost importance in deciding the ability of individual animals to succumb to or resist the effects of a given chemical/disease.

How efficient are differential expression techniques at recovering a high percentage of differentially expressed genes?

A number of groups have produced experimental data suggesting that mammalian cells produce between 8000–15000 different mRNA species at any one time (Mechler and Rabbitts 1981, Hedrick *et al.* 1984, Bravo 1990), although figures as high as 20–30000 have also been quoted (Axel *et al.* 1976). Hedrick *et al.* (1984) provided evidence suggesting that the majority of these belong to the rare abundance class. A breakdown of this abundance distribution is shown in table 1.

When the results of differential display experiments have been compared with data obtained previously using other methods, it is apparent that not all differentially expressed mRNAs are represented in the final display. In particular, rare messages (which, importantly, often include regulatory proteins) are not easily recovered using differential display systems. This is a major shortcoming, as the majority of mRNA species exist at levels of less than 0.005% of the total population (table 1). Bertioli *et al.* (1995) examined the efficiency of DD templates (heterogeneous mRNA populations) for recovering rare messages and were unable to detect mRNA

species present at less than 1.2% of the total mRNA population—equivalent to an intermediate or abundant species. Interestingly, when simple model systems (single target only) were used instead of a heterogeneous mRNA population, the same primers could detect levels of target mRNA down to 10000 × smaller. These results are probably best explained by competition for substrates from the many PCR products produced in a DD reaction.

The numbers of differentially expressed mRNAs reported in the literature using various model systems provides further evidence that many differentially expressed mRNAs are not recovered. For example, DeRisi *et al.* (1997) used DNA array technology to examine gene expression in yeast following exhaustion of sugar in the medium, and found that more than 1700 genes showed a change in expression of at least 2-fold. In light of such a finding, it would not be unreasonable to suggest that of the 8000–15 000 different mRNA species produced by any given mammalian cell, up to 1000 or more may show altered expression following chemical stimulation. Whilst this may be an extreme figure, it is known that at least 100 genes are activated/upregulated in Jurkat (T-) cells following IL-2 stimulation (Ullman *et al.* 1990). In addition, Wan *et al.* (1996) estimated that interferon-γ-stimulated HeLa cells differentially express up to 433 genes (assuming 24000 distinct mRNAs expressed by the cells). However, there have been few publications documenting anywhere near the recovery of these numbers. For example, in using DD to compare normal and regenerating mouse liver, Bauer *et al.* (1993) found only 70 of 38000 total bands to be different. Of these, 50% (35 genes) were shown to correspond to differentially expressed bands. Chen *et al.* (1996) reported 10 genes upregulated in female rat liver following ethinyl estradiol treatment. McKenzie and Drake (1997) identified 14 different gene products whose expression was altered by phorbol myristate acetate (PMA, a tumour promoter agent) stimulation of a human myelomonocytic cell line. Kilty and Vickers (1997) identified 10 different gene products whose expression was upregulated in the peripheral blood leukocytes of allergic disease sufferers. Linskens *et al.* (1995) found 23 genes differentially expressed between young and senescent fibroblasts. Techniques other than DD have also provided an apparent paucity of differentially expressed genes. Using SH for example, Cao *et al.* (1997) found 15 genes differentially expressed in colorectal cancer compared to normal mucosal epithelium. Fitzpatrick *et al.* (1995) isolated 17 genes upregulated in rat liver following treatment with the peroxisome proliferator, clofibrate; Philips *et al.* (1990) isolated 12 cDNA clones which were upregulated in highly metastatic mammary adenocarcinoma cell lines compared to poorly metastatic ones. Prashar and Weissman (1996) used 3′ restriction fragment analysis and identified approximately 40 genes showing altered expression within 4 h of activation of Jurkat T-cells. Groenink and Leegwater (1996) analysed 27 gene fragments isolated using SSH of delayed early response phase of liver regeneration and found only 12 to be upregulated.

In the laboratory, SSH was used to isolate up to 70 candidate genes which appear to show altered expression in guinea pig liver following short-term treatment with the peroxisome proliferator, WY-14,643 (Rockett, Swales, Esdaile and Gibson, unpublished observations). However, these findings have still to be confirmed by analysis of the extracted tissue mRNA for differential expression of these sequences. Whilst the latest differential display technologies are purported to include design and experimental modifications to overcome this lack of efficiency (in both the total number of differentially expressed genes recovered and the percentage that are true

positives), it is still not clear if such adaptations are practically effective—proving efficiency by spiking with a known amount of limited numbers of artificial construct(s) is one thing, but isolating a high percentage of the rare messages already present in an mRNA population is another. Of course, some models will genuinely produce only a small number of differentially expressed genes. In addition, there are also technical problems that can reduce efficiency. For example, mRNAs may have an unusual primary structure that effectively prevents their amplification by PCR-based systems. In addition, it is known that under certain circumstances not all mRNAs have 3′ polyA sites. For example, during *Xenopus* development, deadenylation is used as a means to stabilize RNAs (Voeltz and Steitz 1998), whilst preferential deadenylation may play a role in regulating Hsp70 (and perhaps, therefore, other stress protein) expression in *Drosophila* (Dellavalle *et al.* 1994). The presence of deadenylated mRNAs would clearly reduce the efficiency of systems utilizing a polydT reverse transcription step. The efficiency of any system also depends on the quality of the starting material. All differential display techniques use mRNA as their target material. However, it is difficult to isolate mRNA that is completely free of ribosomal RNA. Even if polydT primers are used to prime first strand cDNA synthesis, ribosomal RNA is often transcribed to some degree (Clontech PCR-Select cDNA Subtraction kit user manual). It has been shown, at least in the case of SSH, that a high rRNA:mRNA ratio can lead to inefficient subtractive hybridization (Clontech PCR-Select cDNA Subtraction kit user manual), and there is no reason to suppose that it will not do likewise in other SH approaches. Finally, those techniques that utilise a presubtraction amplification step (e.g. RDA) may present a skewed representation since some sequences amplify better than others.

Of course, probably the most important consideration is the temporal factor. It is clear that any given differential display experiment can only interrogate a cell at one point in time. It may well be that a high percentage of the genes showing altered expression at that time are obtained. However, given that disease processes and responses to environmental stimuli involve dynamic cascades of signalling, regulation, production and action, it is clear that all those genes which are switched on/off at different times will not be recovered and, therefore, vital information may well be missed. It is, therefore, imperative to obtain as much information about the model system beforehand as possible, from which a strategy can be derived for targeting specific time points or events that are of particular interest to the investigator. One way of getting round this problem of single time point analysis is to conduct the experiment over a suitable time course which, of course, adds substantially to the amount of work involved.

How sensitive are differential expression technologies?

There has been little published data that addresses the issue of how large the change in expression must be for it to permit isolation of the gene in question with the various differential expression technologies. Although the isolation of genes whose expression is changed as little as 1.5-fold has been reported using SSH (Groenink and Leegwater 1996), it appears that those demonstrating a change in excess of 5-fold are more likely to be picked up. Thus, there is a 'grey zone' in between where small changes could fade in and out of isolation between

experiments and animals. DD, on the other hand, is not subject to this grey zone since, unlike SH approaches, it does not amplify the difference in expression between two samples. Wan *et al.* (1996) reported that differences in expression of twofold or more are detectable using DD.

Resolution and visualization of differential expression products

It seems highly improbable with current technology that a gel system could be developed that is able to resolve all gene species showing altered expression in any given test system (be it SH- or DD-based). Polyacrylamide gel electrophoresis (PAGE) can resolve size differences down to 0.2 % (Sambrook *et al.* 1989) and are used as standard in DD experiments. Even so, it is clear that a complex series of gene products such as those seen in a DD will contain unresolvable components. Thus, what appears to be one band in a gel may in fact turn out to be several. Indeed, it has been well documented (Mathieu-Daude *et al.* 1996, Smith *et al.* 1997) that a single band extracted from a DD often represents a composite of heterogeneous products, and the same has been found for SSH displays in this laboratory (Rockett *et al.* 1997). One possible solution was offered by Mathieu-Daude *et al.* (1996), who extracted and reamplified candidate bands from a DD display and used single strand conformation polymorphism (SSCP) analysis to confirm which components represented the truly differentially expressed product.

Many scientists often try to avoid the use of PAGE where possible because it is technically more demanding than agarose gel electrophoresis (AGE). Unfortunately, high resolution agarose gels such as Metaphor (FMC, Lichfield, UK) and AquaPor HR (National Diagnostics, Hessle, UK), whilst easier to prepare and manipulate than PAGE, can only separate DNA sequences which differ in size by around 1.5–2 % (15–20 base pairs for a 1Kb fragment). Thus, SSH, RDA or other such products which differ in size by less than this amount are normally not resolvable. However, a simple technique does in fact exist for increasing the resolving power of AGE—the inclusion of HA-red (10-phenyl neutral red-PEG ligand) or HA-yellow (bisbenzamide-PEG ligand) (Hanse Analytik GmbH, Bremen, Germany) in a gel separates identical or closely sized products on base content. Specifically, HA-red and -yellow selectively bind to GC and AT DNA motifs, respectively (Wawer *et al.* 1995, Hanse Analytik 1997, personal communication). Since both HA-stains possess an overall positive charge, they migrate towards the cathode when an electric field is applied. This is in direct opposition to DNA, which is negatively charged and, therefore, migrates towards the anode. Thus, if two DNA clones are identical in size (as perceived on a standard high resolution agarose gel), but differ in AT/GC content, inclusion of a HA-dye in the gel will effectively retard the migration of one of the sequences compared to the other, effectively making it apparently larger and, thus, providing a means of differentiating between the two. The use of HA-red has been shown to resolve sequences with an AT variation of less than 1 % (Wawer *et al.* 1995), whilst Hanse Analytik have reported that HA staining is so sensitive that in one case it was used to distinguish two 567bp sequences which differed by only a single point mutation (Hanse Analytik 1996, personal communication). Therefore, if one wishes to check whether all the clones produced from a specific band in a differential display experiment are derived from the same gene species, a small amount of reamplified or digested clone can be run on a standard high resolution gel, and a second aliquot

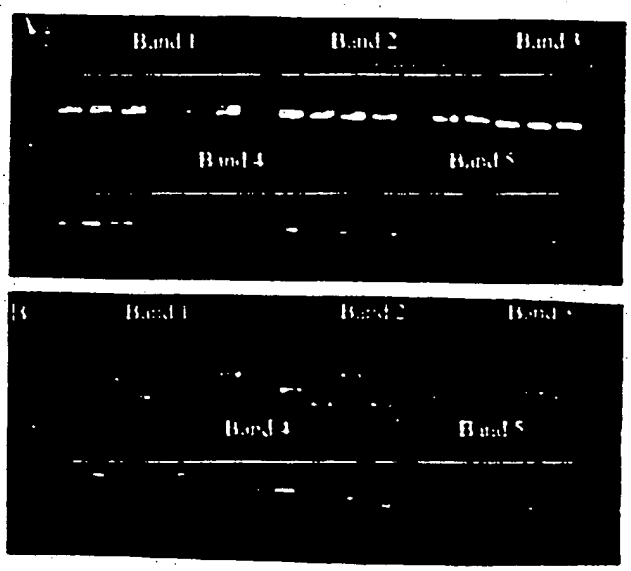


Figure 10. Discrimination of clones of identical/nearly identical size using HA-red. Bands of decreasing size (1–5) were extracted from the final display of a suppression subtractive hybridization experiment and cloned. Seven colonies were picked at random from each cloned band and their inserts amplified using PCR. The products were run on two gels, (A) a high resolution 2% agarose gel, and (B) a high resolution 2% agarose gel containing 1 U/ml HA-red. With few exceptions, all the clones from each band appear to be the same size (gel A). However, the presence of HA-red (gel B), which separates identically-sized DNA fragments based on the percentage of GC within the sequence, clearly indicates the presence of different gene species within each band. For example, even though all five re-amplified clones of band 1 appear to be the same size, at least four different gene species are represented.

in a similar gel containing one of the HA-stains. The standard gel should indicate any gross size differences, whilst the HA-stained gel should separate otherwise unresolvable species (on standard AGE) according to their base content. Geisinger *et al.* (1997) reported successful use of this approach for identifying DD-derived clones. Figure 10 shows such an experiment carried out in this laboratory on clones obtained from a band extracted from an SSH display.

An alternative approach is to carry out a 2-D analysis of the differential display products. In this approach, size-based separation is first carried out in a standard agarose gel. The gel slice containing the display is then extracted and incorporated in to a HA gel for resolution based on AT/GC content.

Of course, one should always consider the possibility of there being different gene species which are the same size and have the same GC/AT content. However, even these species are not unresolvable given some effort—again, one might use SSCP, or perhaps a denaturing gradient gel electrophoresis (DGGE) or temperature gradient field electrophoresis (TGGE) approach to resolve the contents of a band, either directly on the extracted band (Suzuki *et al.* 1991) or on the reamplified product.

The requirement of some differential display techniques to visualize large numbers of products (e.g. DD and GEF) can also present a problem in that, in terms of numbers, the resolution of PAGE rarely exceeds 300–400 bands. One approach to overcoming this might be to use 2-D gels such as those described by Uitterlinden *et al.* (1989) and Hatada *et al.* (1991).

Extraction of differentially expressed bands from a gel can be complex since, in some cases (e.g. DD, GEF), the results are visualized by autoradiographic means, such that precise overlay of the developed film on the gel must occur if the correct band is to be extracted for further analysis. Clearly, a misjudged extraction can account for many man-hours lost. This problem, and that of the use of radioisotopes, has been addressed by several groups. For example, Lohmann *et al.* (1995) demonstrated that silver staining can be used directly to visualize DD bands in horizontal PAGs. An *et al.* (1996) avoided the use of radioisotopes by transferring a small amount (20–30%) of the DNA from their DD to a nylon membrane, and visualizing the bands using chemiluminescent staining before going back to extract the remaining DNA from the gel. Chen and Peck (1996) went one step further and transferred the entire DD to a nylon membrane. The DNA bands were then visualized using a digoxigenin (DIG) system (DIG was attached to the polydT primers used in the differential display procedure). Differentially expressed bands were cut from the membrane and the DNA eluted by washing with PCR buffer prior to reamplification.

One of the advantages of using techniques such as SSH and RDA is that the final display can be run on an agarose gel and the bands visualized with simple ethidium bromide staining. Whilst this approach can provide acceptable results, overstaining with SYBR Green I or SYBR Gold nucleic acid stains (FMC) effectively enhances the intensity and sharpness of the bands. This greatly aids in their precise extraction and often reveals some faint products that may otherwise be overlooked. Whilst differential displays stained with SYBR Green I are better visualized using short wavelength UV (254 nm) rather than medium wavelength (306 nm), the shorter wavelength is much more DNA damaging. In practice, it takes only a few seconds to damage DNA extracted under 254 nm irradiation, effectively preventing reamplification and cloning. The best approach is to overstain with SYBR Green I and extract bands under a medium wavelength UV transillumination.

The possible use of 'microfingerprinting' to reduce complexity

Given the sheer number of gene products and the possible complexity of each band, an alternative approach to rapid characterization may be to use an enhanced analysis of a small section of a differential display—a 'sub-fingerprint' or 'micro-fingerprint'. In this case, one could concentrate on those bands which only appear in a particular chosen size region. Reducing the fingerprint in this way has at least two advantages. One is that it should be possible to use different gel types, concentrations and run times tailored exactly to that region. Currently, one might run products from 100–3000+ bp on the same gel, which leads to compromize in the gel system being used and consequently to suboptimal resolution, both in terms of size and numbers, and can lead to problems in the accurate excision of individual bands. Secondly, it may be possible to enhance resolution by using a 2-D analysis using a HA-stain, as described earlier. In summary, if a range of gene product sizes is carefully chosen to included certain 'relevant' genes, the 2-D system standardized, and appropriate gene analysis used, it may be possible to develop a method for the early and rapid identification of compounds which have similar or widely different cellular effects. If the prognosis for exposure to one or more other chemicals which display a similar profile is already known, then one could perhaps predict similar effects for any new compounds which show a similar micro-fingerprint.

An alternative approach to microfingerprinting is to examine altered expression in specific families of genes through careful selection of PCR primers and/or post-reaction analysis. Stress genes, growth factors and/or their receptors, cell cycling genes, cytochromes P450 and regulatory proteins might be considered as candidates for analysis in this way. Indeed, some off-the-shelf DNA arrays (e.g. Clontech's Atlas cDNA Expression Array series) already anticipated this to some degree by grouping together genes involved in different responses e.g. apoptosis, stress, DNA-damage response etc.

Screening

False positives

The generation of false positives has been discussed at length amongst the differential display community (Liang *et al.* 1993, 1995, Nishio *et al.* 1994, Sun *et al.* 1994, Sompayrac *et al.* 1995). The reason for false positives varies with the technique being used. For instance, in RDA, the use of adaptors which have not been HPLC purified can lead to the production of false positives through illegitimate ligation events (O'Neill and Sinclair 1997), whilst in DD they can arise through PCR artifacts and illegitemate transcription of rRNA. In SH, false positives appear to be derived largely from abundant gene species, although some may arise from cDNA/mRNA species which do not undergo hybridization for technical reasons.

A quick screening of putative differentially expressed clones can be carried out using a simple dot blot approach, in which labelled first strand probes synthesized from tester and driver mRNA are hybridized to an array of said clones (Hedrick *et al.* 1984, Sakaguchi *et al.* 1986). Differentially expressed clones will hybridize to tester probe, but not driver. The disadvantage of this approach is that rare species may not generate detectable hybridization signals. One option for those using SSH is to screen the clones using a labelled probe generated from the subtracted cDNA from which it was derived, and with a probe made from the reverse subtraction reaction (ClonTechniques 1997a). Since the SSH method enriches rare sequences, it should be possible to confirm the presence of clones representing low abundance genes. Despite this quick screening step, there is still the need to go back to the original mRNA and confirm the altered expression using a more quantitative approach. Although this may be achieved using Northern blots, the sensitivity is poor by today's high standards and one must rely on PCR methods for accurate and sensitive determinations (see below).

Sequence analysis

The majority of differential display procedures produce final products which are between 100 and 1000bp in size. However, this may considerably reduce the size of the sequence for analysis of the DNA databases. This in turn leads to a reduced confidence in the result—several families of genes have members whose DNA sequences are almost identical except in a few key stretches, e.g. the cytochrome P450 gene superfamily (Nelson *et al.* 1996). Thus, does the clone identified as being almost identical to gene X_0 really come from that gene, or its brother gene X_1 or its as yet undiscovered sister X_2? For example, using SSH, part of a gene was isolated,

which was up-regulated in the liver of rats exposed to Wy-14,643 and was identified by a FASTA search as being transferrin (data not shown). However, transferrin is known to be downregulated by hypolipidemic peroxisome proliferators such as Wy-14,643 (Hertz *et al.* 1996), and this was confirmed with subsequent RT-PCR analysis. This suggests that the gene sequence isolated may belong to a gene which is closely related to transferrin, but is regulated by a different mechanism.

A further problem associated with SH technology is redundancy. In most cases before SH is carried out, the cDNA population must first be simplified by restriction digestion. This is important for at least two reasons:

(1) To reduce complexity—long cDNA fragments may form complex networks which prevent the formation of appropriate hybrids, especially at the high concentrations required for efficient hybridization.
(2) Cutting the cDNAs into small fragments provides better representation of individual genes. This is because genes derived from related but distinct members of gene families often have similar coding sequences that may cross-hybridize and be eliminated during the subtraction procedure (Ko 1990). Furthermore, different fragments from the same cDNA may differ considerably in terms of hybridization and amplification and, thus, may not efficiently do one or the other (Wang and Brown 1991). Thus, some fragments from differentially expressed cDNAs may be eliminated during subtractive hybridization procedures. However, other fragments may be enriched and isolated. As a consequence of this, some genes will be cut one or more times, giving rise to two or more fragments of different sizes. If those same genes are differentially expressed, then two or more of the different size fragments may come through as separate bands on the final differential display, increasing the observed redundancy and increasing the number of redundant sequencing reactions.

Sequence comparisons also throw up another important point—at what degree of sequence similarity does one accept a result. Is 90% identitiy between a gene derived from your model species and another acceptably close? Is 95% between your sequence and one from the same species also acceptable? This problem is particularly relevant when the forward and reverse sequence comparisons give similar sequences with completely different gene species! An arbitrary decision seems to be to allocate genes that are definite (95% and above similarity) and then group those between 60 and 95% as being related or possible homologues.

Quantitative analysis

At some point, one must give consideration to the quantitative analysis of the candidate genes, either as a means of confirming that they are truly differentially expressed, or in order to establish just what the differences are. Northern blot analysis is a popular approach as it is relatively easy and quick to perform. However, the major drawback with Northern blots is that they are often not sensitive enough to detect rare sequences. Since the majority of messages expressed in a cell are of low abundance (see table 1), this is a major problem. Consequently, RT-PCR may be the method of choice for confirming differential expression. Although the procedure is somewhat more complex than Northern analysis, requiring synthesis of primers and optimization of reaction conditions for each gene species, it is now possible to set up high throughput PCR systems using mulitchannel pipettes, 96+-well plates and

appropriate thermal cycling technology. Whilst quantitative analysis is more desirable, being more accurate and without reliance on an internal standard, the money and time needed to develop a competitor molecule is often excessive, especially when one might be examining tens or even hundreds of gene species. The use of semi-quantitative analysis is simpler, although still relatively involved. One must first of all choose an internal standard that does not change in the test cells compared to the controls. Numerous reference genes have been tried in the past, for example interferon-gamma (IFN-γ, Frye *et al.* 1989), β-actin (Heuval *et al.* 1994), glyceraldehyde-3-phosphate dehydrogenase (GAPDH, Wong *et al.* 1994), dihydrofolate reductase (DHFR, Mohler and Butler 1991), β-2-microglobulin (β-2-m, Murphy *et al.* 1990), hypoxanthine phosphoribosyl transferase (HPRT, Foss *et al.* 1998) and a number of others (ClonTechniques 1997b). Ideally, an internal standard should not change its level of expression in the cell regardless of cell age, stage in the cell cycle or through the effects of external stimuli. However, it has been shown on numerous occasions that the levels of most housekeeping genes currently used by the research community do in fact change under certain conditions and in different tissues (ClonTechniques 1997b). It is imperative, therefore, that preliminary experiments be carried out on a panel of housekeeping genes to establish their suitability for use in the model system.

Interpretation of quantitative data must also be treated with caution. By comparing the lists of genes identified by differential expression one can perhaps gain insight into why two different species react in different ways to external stimuli. For example, rats and mice appear sensitive to the non-genotoxic effects of a wide range of peroxisome proliferators whilst Syrian hamsters and guinea pigs are largely resistant (Orton *et al.* 1984, Rodricks and Turnbull 1987, Lake *et al.* 1989, 1993, Makowska *et al.* 1992). A simplified approach to resolving the reason(s) why is to compare lists of up- and down-regulated genes in order to identify those which are expressed in only one species and, through background knowledge of the effects of the said gene, might suggest a mechanism of facilitated non-genotoxic carcinogenesis or protection. Of course, the situation is likely to be far more complex. Perhaps if there were one key gene protecting guinea pig from non-genotoxic effects and it was upregulated 50 times by PPs, the same gene might only be up-regulated five times in the rat. However, since both were noted to be upregulated, the importance of the gene may be overlooked. Just to complicate matters, a large change in expression does not necessarily mean a biologically important change. For example, what is the true relevance of gene Y which shows a 50-fold increase after a particular treatment, and gene Z which shows only a 5-fold increase? If one examines the literature one may find that historically, gene Y has often been shown to be up-regulated 40–60-fold by a number of unrelated stimuli—in light of this the 50-fold increase would appear less significant. However, the literature may show that gene Z has never been recorded as having more than doubled in expression—which makes your 5-fold increase all the more exciting. Perhaps even more interesting is if that same 5-fold increase has only been seen in related neoplasms or following treatment with related chemicals.

Problems in using the differential display approach

Differential display technology originally held promise of an easily obtainable 'fingerprint' of those genes which are up- or down-regulated in test animals/cells in a developmental process or following exposure to given stimuli. However, it has

become clear that the fingerprinting process, whilst still valid, is much too complex to be represented by a single technique profile. This is because all differential display techniques have common and/or unique technical problems which preclude the isolation and identification of all those genes which show changes in expression. Furthermore, there are important genetic changes related to disease development which differential expression analysis is simply not designed to address. An example of this is the presence of small deletions, insertions, or point mutations such as those seen in activated oncogenes, tumour suppressor genes and individual polymorphisms. Polymorphic variations, small though they usually are, are often regarded as being of paramount importance in explaining why some patients respond better than others to certain drug treatments (and, in logical extension, why some people are less affected by potentially dangerous xenobiotics/carcinogens than others). The identification of such point mutations and naturally occurring polymorphisms requires the subsequent application of sequencing, SSCP, DGGE or TGGE to the gene of interest. Furthermore, differential display is not designed to address issues such as alternatively spliced gene species or whether an increased abundance of mRNA is a result of increased transcription or increased mRNA stability.

Conclusions

Perhaps the main advantage of open system differential display techniques is that they are not limited by extant theories or researcher bias in revealing genes which are differentially expressed, since they are designed to amplify all genes which demonstrate altered expression. This means that they are useful for the isolation of previously unknown genes which may turn out be useful biomarkers of a particular state or condition. At least one open system (SAGE) is also quantitative, thus eliminating the need to return to the original mRNA and carry out Northern/PCR analysis to confirm the result. However, the rapid progress of genome mapping projects means that over the next 5–10 years or so, the balance of experimental use will switch from open to closed differential display systems, particularly DNA arrays. Arrays are easier and faster to prepare and use, provide quantitative data, are suitable for high throughput analysis and can be tailored to look at specific signalling pathways or families of genes. Identification of all the gene sequences in human and common laboratory animals combined with improved DNA array technology, means that it will soon no longer be necessary to try to isolate differentially expressed genes using the technically more demanding open system approach. Thus, their main advantage (that of identifying unknown genes) will be largely eradicated. It is likely, therefore, that their sphere of application will be reduced to analysis of the less common laboratory species, since it will be some time yet before the genomes of such animals as zebrafish, electric eels, gerbils, crayfish and squid, for example, will be sequenced.

Of course, in the end the question will always remain: What is the functional/biological significance of the identified, differentially expressed genes? One persistent problem is understanding whether differentially expressed genes are a cause or consequence of the altered state. Furthermore, many chemicals, such as non-genotoxic carcinogens, are also mitogens and so genes associated with replication will also be upregulated but may have little or nothing to do with the

carcinogenic effect. Whilst differential display technology cannot hope to answer these questions, it does provide a springboard from which identification, regulatory and functional studies can be launched. Understanding the molecular mechanism of cellular responses is almost impossible without knowing the regulation and function of those genes and their condition (e.g. mutated). In an abstract sense, differential display can be likened to a still photograph, showing details of a fixed moment in time. Consider the Historian who knows the outcome of a battle and the placement and condition of the troops before the battle commenced, but is asked to try and deduce how the battle progressed and why it ended as it did from a few still photographs—an impossible task. In order to understand the battle, the Historian must find out the capabilities and motivation of the soldiers and their commanding officers, what the orders were and whether they were obeyed. He must examine the terrain, the remains of the battle and consider the effects the prevailing weather conditions exerted. Likewise, if mechanistic answers are to be forthcoming, the scientist must use differential display in combination with other techniques, such as knockout technology, the analysis of cell signalling pathways, mutation analysis and time and dose response analyses. Although this review has emphasized the importance of differential gene profiling, it should not be considered in isolation and the full impact of this approach will be strengthened if used in combination with functional genomics and proteomics (2-dimensional protein gels from isoelectric focusing and subsequent SDS electrophoresis and virtual 2D-maps using capillary electrophoresis). Proteomics is attracting much recent attention as many of the changes resulting in differential gene expression do not involve changes in mRNA levels, as decribed extensively herein, but rather protein–protein, protein–DNA and protein phosphorylation events which would require functional genomics or proteomic technologies for investigation.

Despite the limitations of differential display technology, it is clear that many potential applications and benefits can be obtained from characterizing the genetic changes that occur in a cell during normal and disease development and in response to chemical or biological insult. In light of functional data, such profiling will provide a 'fingerprint' of each stage of development or response, and in the long term should help in the elucidation of specific and sensitive biomarkers for different types of chemical/biological exposure and disease states. The potential medical and therapeutic benefits of understanding such molecular changes are almost immeasurable. Amongst other things, such fingerprints could indicate the family or even specific type of chemical an individual has been exposed to plus the length and/or acuteness of that exposure, thus indicating the most prudent treatment. They may also help uncover differences in histologically identical cancers, provide diagnostic tests for the earliest stages of neoplasia and, again, perhaps indicate the most efficacious treatment.

The Human Genome Project will be completed early in the next century and the DNA sequence of all the human genes will be known. The continuing development and evolution of differential gene expression technology will ensure that this knowledge contributes fully to the understanding of human disease processes.

Acknowledgements

We acknowledge Drs Nick Plant (University of Surrey), Sally Darney and Chris Luft (US EPA at RTP) for their critical analysis of the manuscript prior to submission. This manuscript has been reviewed in accordance with the policy of the

US Environmental Protection Agency and approved for publication. Approval does not signify that the contents reflect the views and policies of the Agency, nor does mention of trade names constitute endorsement or recommendation for use.

References

ADAMS, M. D., KELLEY, J. M., GOCAYNE, J. D., DUBNICK, M., POLYMEROPOULOS, M. H., XIAO, H., MERRIL, C. R., WU, A., OLDE, B., MORENO, R. F., KERLAVAGE, A. R., MCCOMBIE, W. R. and VENTOR, J. C., 1991, Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, **252**, 1651–1656.

AN, G., LUO, G., VELTRI, R. W. and O'HARA, S. M., 1996, Sensitive non-radioactive differential display method using chemiluminescent detection. *Biotechniques*, **20**, 342–346.

AXEL, R., FEIGELSON, P. and SCHULTZ, G., 1976, Analysis of the complexity and diversity of mRNA from chicken liver and oviduct. *Cell*, **7**, 247–254.

BAND, V. and SAGER, R., 1989, Distinctive traits of normal and tumor-derived human mammary epithelial cells expressed in a medium that supports long-term growth of both cell types. *Proceedings of the National Academy of Sciences, USA*, **86**, 1249–1253.

BAUER, D., MULLER, H., REICH, J., RIEDEL, H., AHRENKIEL, V., WARTHOE, P. and STRAUSS, M., 1993, Identification of differentially expressed mRNA species by an improved display technique (DDRT-PCR). *Nucleic Acids Research*, **21**, 4272–4280.

BERTIOLI, D. J., SCHLICHTER, U. H. A., ADAMS, M. J., BURROWS, P. R., STEINBISS, H.-H. and ANTONIW, J. F., 1995, An analysis of differential display shows a strong bias towards high copy number mRNAs. *Nucleic Acids Research*, **23**, 4520–4523.

BRAVO, R., 1990, Genes induced during the G0/G1 transition in mouse fibroblasts. *Seminars in Cancer Biology*, **1**, 37–46.

BURN, T. C., PETROVICK, M. S., HOHAUS, S., ROLLINS, B. J. and TENEN, D. G., 1994, Monocyte chemoattractant protein-1 gene is expressed in activated neutrophils and retinoic acid-induced human myeloid cell lines. *Blood*, **84**, 2776–2783.

CAO, J., CAI, X., ZHENG, L., GENG, L., SHI, Z., PAO, C. C. and ZHENG, S., 1997, Characterisation of colorectal cancer-related cDNA clones obtained by subtractive hybridisation screening. *Journal of Cancer Research and Clinical Oncology*, **123**, 447–451.

CASSIDY, S. B., 1995, Uniparental disomy and genomic imprinting as causes of human genetic disease. *Environmental and Molecular Mutagenesis*, **25** (Suppl 26), 13–20.

CHANG, G. W. and TERZAGHI-HOWE, M., 1998, Multiple changes in gene expression are associated with normal cell-induced modulation of the neoplastic phenotype. *Cancer Research*, **58**, 4445–4452.

CHEN, J., SCHWARTZ, D. A., YOUNG, T. A., NORRIS, J. S. and YAGER, J. D., 1996, Identification of genes whose expression is altered during mitosuppression in livers of ethinyl estradiol-treated female rats. *Carcinogenesis*, **17**, 2783–2786.

CHEN, J. J. W. and PECK, K., 1996, Non-radioactive differential display method to directly visualise and amplify differential bands on nylon membrane. *Nucleic Acid Research*, **24**, 793–794.

CLONTECHNIQUES, 1997a, PCR-Select Differential Screening Kit—the nextstep after Clontech PCR-Select cDNA subtraction. *ClonTechniques*, **XII**, 18–19.

CLONTECHNIQUES, 1997b, Housekeeping RT-PCR amplimers and cDNA probes. *ClonTechniques*, **XII**, 15–16.

DAVIS, M. M., COHEN, D. I., NIELSEN, E. A., STEINMETZ, M., PAUL, W. E. and HOOD, L., 1984, Cell-type-specific cDNA probes and the murine I region: the localization and orientation of Ad alpha. *Proceedings of the National Academy of Sciences (USA)*, **81**, 2194–2198.

DELLAVALLE, R. P., PETERSON, R. and LINDQUIST, S., 1994, Preferential deadenylation of HSP70 mRNA plays a key role in regulating Hsp70 expression in Drosophila melanogaster. *Molecular and Cell Biology*, **14**, 3646–3659.

DERISI, J. L., VASHWANATH, R. L. and BROWN, P., 1997, Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**, 680–686.

DIATCHENKO, L., LAU, Y.-F. C., CAMPBELL, A. P., CHENCHIK, A., MOQADAM, F., HUANG, B., LUKYANOV, K., GURSKAYA, N., SVERDLOV, E. D. and SIEBERT, P. D., 1996, Suppression subtractive hybridisation: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proceedings of the National Academy of Sciences (USA)*, **93**, 6025–6030.

DOGRA, S. C., WHITELAW, M. L. and MAY, B. K., 1998, Transcriptional activation of cytochrome P450 genes by different classes of chemical inducers. *Clinical and Experimental Pharmacology and Physiology*, **25**, 1–9.

DUGUID, J. R. and DINAUER, M. C., 1990, Library subtraction of *in vitro* cDNA libraries to identify differentially expressed genes in scrapie infection. *Nucleic Acids Research*, **18**, 2789–2792.

DUNBAR, P. R., OGG, G. S., CHEN, J., RUST, N., VAN DER BRUGGEN, P. and CERUNDOLO, V., 1998, Direct isolation, phenotyping and cloning of low-frequency antigen-specific cytotoxic T lymphocytes from peripheral blood. *Current Biology*, **26**, 413–416.

FITZPATRICK, D. R., GERMAIN-LEE, E. and VALLE, D., 1995, Isolation and characterisation of rat and human cDNAs encoding a novel putative peroxisomal enoyl-CoA hydratase. *Genomics*, **27**, 457–466.

FOSS, D. L., BAARSCH, M. J. and MURTAUGH, M. P., 1998, Regulation of hypoxanthine phosphoribosyltransferase, glyceraldehyde-3-phosphate dehydrogenase and beta-actin mRNA expression in porcine immune cells and tissues. *Animal Biotechnology*, **9**, 67–78 .

FRYE, R. A., BENZ, C. C. and LIU, E., 1989, Detection of amplified oncogenes by differential polymerase chain reaction. *Oncogene*, **4**, 1153–1157.

GEISINGER, A., RODRIGUEZ, R., ROMERO, V. and WETTSTEIN R., 1997, A simple method for screening cDNAs arising from the cloning of RNA differential display bands. *Elsevier Trends Journals Technical Tips Online*, http://tto.trends.com, document T01110.

GRESS, T. M., HOHEISEL, J. D., LENNON, G. G., ZEHETNER, G. and LEHRACH, H., 1992, Hybridisation fingerprinting of high density cDNA filter arrays with cDNA pools derived from whole tissues. *Mammalian Genome*, **3**, 609–619.

GRIFFIN, G. and KRISHNA, S., 1998, Cytokines in infectious diseases. *Journal of the Royal College of Physicians, London*, **32**, 195–198.

GROENINK, M. and LEEGWATER, A. C. J., 1996, Isolation of delayed early genes associated with liver regeneration using Clontech PCR-select subtraction technique. *Clontechniques*, **XI**, 23–24.

GUIMARAES, M. J., BAZAN, J. F., ZLOTNIK, A., WILES, M. V., GRIMALDI, J. C., LEE, F. and MCCLANAHAN, T., 1995b, A new approach to the study of haematopoietic development in the yolk sac and embryoid bodies. *Development*, **121**, 3335–3346.

GUIMERAES, M. J., LEE, F., ZLOTNIK, A. and MCCLANAHAN, T., 1995a, Differential display by PCR: novel findings and applications. *Nucleic Acids Research*, **23**, 1832–1833.

GURSKAYA, N. G., DIATCHENKO, L., CHENCHIK, P. D., SIEBERT, P. D., KHASPEKOV, G. L., LUKYANOV, K. A., VAGNER, L. L., ERMOLAEVA, O. D., LUKYANOV, S. A. and SVERDLOV, E. D., 1996, Equalising cDNA subtraction based on selective suppression of polymerase chain reaction: Cloning of Jurkat cell transcripts induced by phytohemaglutinin and phorbol 12-Myrystate 13-Acetate. *Analytical Biochemistry*, **240**, 90–97.

HAMPSON, I. N. and HAMPSON, L., 1997, CCLS and DROP—subtractive cloning made easy. *Life Science News* (A publication of Amersham Life Science), **23**, 22–24.

HAMPSON, I. N., HAMPSON, L. and DEXTER, T. M., 1996, Directional random oligonucleotide primed (DROP) global amplification of cDNA: its application to subtractive cDNA cloning. *Nucleic Acids Research*, **24**, 4832–4835.

HAMPSON, I. N., POPE, L., COWLING, G. J. and DEXTER, T. M., 1992, Chemical cross linking subtraction (CCLS): a new method for the generation of subtractive hybridisation probes. *Nucleic Acids Research*, **20**, 2899.

HARA, E., KATO, T., NAKADA, S., SEKIYA, S. and ODA, K., 1991, Subtractive cDNA cloning using oligo(dT)30-latex and PCR: isolation of cDNA clones specific to undifferentiated human embryonal carcinoma cells. *Nucleic Acids Research*, **19**, 7097–7104.

HATADA, I., HAYASHIZAKE, Y., HIROTSUNE, S., KOMATSUBARA, H. and MUKAI, T., 1991, A genomic scanning method for higher organisms using restriction sites as landmarks. *Proceedings of the National Academy of Sciences (USA)*, **88**, 9523–9527.

HECHT, N., 1998, Molecular mechanisms of male sperm cell differentiation. *Bioessays*, **20**, 555–561.

HEDRICK, S., COHEN, D. I., NIELSEN, E. A. and DAVIS, M. E., 1984, Isolation of T cell-specific membrane-associated proteins. *Nature*, **308**, 149–153.

HERTZ, R., SECKBACH, M., ZAKIN, M. M. and BAR-TANA, J., 1996, Transcriptional suppression of the transferrin gene by hypolipidemic peroxisome proliferators. *Journal of Biological Chemistry*, **271**, 218–224.

HEUVAL, J. P. V., CLARK, G. C., KOHN, M. C., TRITSCHER, A. M., GREENLEE, W. F., LUCIER, G. W. and BELL, D. A., 1994, Dioxin-responsive genes: Examination of dose-response relationships using quantitative reverse transciptase-polymerase chain reaction. *Cancer Research*, **54**, 62–68.

HILLIER, L. D., LENNON, G., BECKER, M., BONALDO, M. F., CHIAPELLI, B., CHISSOE, S., DIETRICH, N., DUBUQUE, T., FAVELLO, A., GISH, W., HAWKINS, M., HULTMAN, M., KUCABA, T., LACY, M., LE, M., LE, N., MARDIS, E., MOORE, B., MORRIS, M., PARSONS, J., PRANGE, C., RIFKIN, L., ROHLFING, T., SCHELLENBERG, K., SOARES, M. B., TAN, F., THIERRY-MEG, J., TREVASKIS, E., UNDERWOOD, K., WOHLDMAN, P., WATERSTON, R., WILSON, R and MARRA, M., 1996, Generation and analysis of 280,000 human expressed sequence tags. *Genome Research*, **6**, 807–828.

HUBANK, M. and SCHATZ, D. G., 1994, Identifying differences in mRNA expression by representational difference analysis. *Nucleic Acids Research*, **22**, 5640–5648.

HUNTER, T., 1991, Cooperation between oncogenes. *Cell*, **64**, 249–270.

IVANOVA, N. B. and BELYAVSKY, A. V., 1995, Identification of differentially expressed genes by restriction endonuclease-based gene expression fingerprinting. *Nucleic Acids Research*, **23**, 2954–2958.

JAMES, B. D. and HIGGINS, S. J, 1985, *Nucleic Acid Hybridisation* (Oxford: IRL Press Ltd).

KAS-DEELEN, A. M., HARMSEN, M. C., DE MAAR, E. F. and VAN SON, W. J, 1998, A sensitive method for

quantifying cytomegalic endothelial cells in peripheral blood from cytomegalovirus-infected patients. *Clinical Diagnostic and Laboratory Immunology*, **5**, 622–626.

KILTY, I. and VICKERS, P., 1997, Fractionating DNA fragments generated by differential display PCR. *Strategies Newsletter* (Stratagene), **10**, 50–51.

KLEINJAN, D.-J. and VAN HEYNINGEN, V., 1998, Position effect in human genetic disease. *Human and Molecular Genetics*, **7**, 1611–1618.

KO, M. S., 1990, An 'equalized cDNA library' by the reassociation of short double-stranded cDNAs. *Nucleic Acids Research*, **18**, 5705–5711.

LAKE, B. G., EVANS, J. G., CUNNINGHAME, M. E. and PRICE, R. J., 1993, Comparison of the hepatic effects of Wy-14,643 on peroxisome proliferation and cell replication in the rat and Syrian hamster. *Environmental Health Perspectives*, **101**, 241–248.

LAKE, B. G., EVANS, J. G., GRAY, T. J. B., KOROSI, S. A. and NORTH, C. J., 1989, Comparative studies of nafenopin-induced hepatic peroxisome proliferation in the rat, Syrian hamster, guiea pig and marmoset. *Toxicology and Applied Pharmacology*, **99**, 148–160.

LENNARD, M. S., 1993, Genetically determined adverse drug reactions involving metabolism. *Drug Safety*, **9**, 60–77.

LEVY, S., TODD, S. C. and MAECKER, H. T., 1998, CD81(TAPA-1): a molecule involved in signal transduction and cell adhesion in the immune system. *Annual Review of Immunology*, **16**, 89–109.

LIANG, P. and PARDEE, A. B., 1992, Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, **257**, 967–971.

LIANG, P., AVERBOUKH, L., KEYOMARSI, K., SAGER, R. and PARDEE, A., 1992, Differential display and cloning of messenger RNAs from human breast cancer versus mammary epithelial cells. *Cancer Research*, **52**, 6966–6968.

LIANG, P., AVERBOUKH, L. and PARDEE, A. B., 1993, Distribution & cloning of eukaryotic mRNAs by means of differential display refinements and optimisation. *Nucleic Acids Research*, **21**, 3269–3275.

LIANG, P., BAUER, D., AVERBOUKH, L., WARTHOE, P., ROHRWILD, M., MULLER, H., STRAUSS, M. and PARDEE, A. B., 1995, Analysis of altered gene expression by differential display. *Methods in Enzymology*, **254**, 304–321.

LINSKENS, M. H., FENG, J., ANDREWS, W. H., ENLOW, B. E., SAATI, S. M., TONKIN, L. A., FUNK, W. D. and VILLEPONTEAU, B., 1995, Cataloging altered gene expression in young and senescent cells using enhanced differential display. *Nucleic Acids Research*, **23**, 3244–3251.

LISITSYN, N., LISITSYN, N. and WIGLER, M., 1993, Cloning the differences between two complex genomes. *Science*, **259**, 946–951.

LOHMANN, J., SCHICKLE, H. and BOSCH, T. C. G., 1995, REN Display, a rapid and efficient method for non-radioactive differential display and mRNA isolation. *Biotechniques*, **18**, 200–202.

LUNNEY, J. K., 1998, Cytokines orchestrating the immune response. *Reviews in Science and Techology*, **17**, 84–94.

MAKOWSKA, J. M., GIBSON, G. G. and BONNER, F. W., 1992, Species differences in ciprofibrate-induction of hepaic cytochrome P4504A1 and peroxisome proliferation. *Journal of Biochemical Toxicology*, **7**, 183–191.

MALDARELLI, F., XIANG, C., CHAMOUN, G. and ZEICHNER, S. L., 1998, The expression of the essential nuclear splicing factor SC35 is altered by human immunodeficiency virus infection. *Virus Research*, **53**, 39–51.

MATHIEU-DAUDE, F., CHENG, R., WELSH, J. and MCCLELLAND, M., 1996, Screening of differentially amplified cDNA products from RNA arbitrarily primed PCR fingerprints using single strand conformation polymorphism (SSCP) gels. *Nucleic Acids Research*, **24**, 1504–1507.

MCKENZIE, D. and DRAKE, D., 1997, Identification of differentially expressed gene products with the castaway system. *Strategies Newsletter* (Stratagene), **10**, 19–20.

MCCLELLAND, M., MATHIEU-DAUDE, F. and WELSH, J., 1996, RNA fingerprinting and differential display using arbitrarily primed PCR. *Trends in Genetics*, **11**, 242–246.

MECHLER, B. and RABBITTS, T. H., 1981, Membrane-bound ribosomes of myeloma cells. IV. mRNA complexity of free and membrane-bound polysomes. *Journal of Cell Biology*, **88**, 29–36.

MEYER, U. A. and ZANGER, U. M., 1997, Molecular mechanisms of genetic polymorphisms of drug metabolism. *Annual Review of Pharmacology and Toxicology*, **37**, 269–296.

MOHLER, K. M. and BUTLER, L. D., 1991, Quantitation of cytokine mRNA levels utilizing the reverse transcriptase-polymerase chain reaction following primary antigen-specific sensitization in vivo—I. Verification of linearity, reproducibility and specificity. *Molecular Immunology*, **28**, 437–447.

MURPHY, L. D., HERZOG, C. E., RUDICK, J. B., TITO FOJO, A. and BATES, S. E., 1990, Use of the polymerase chain reaction in the quantitation of the mdr-1 gene expression. *Biochemistry*, **29**, 10351–10356.

NELSON, D. R., KOYMANS, L., KAMATAKI, T., STEGEMAN, J. J., FEYEREISEN, R., WAXMAN, D. J., WATERMAN, M. R., GOTOH, O., COON, M. J., ESTABTROOK, R. W., GUNSALUS, I. C. and NEBERT, D. W., 1996, Update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics*, **6**, 1–42.

NISHIO, Y., AIELLO, L. P. and KING, G. L., 1994, Glucose induced genes in bovine aortic smooth muscle cells identified by mRNA differential display. *FASEB Journal*, **8**, 103–106.

O'NEILL, M. J. and SINCLAIR, A. H., 1997, Isolation of rare transcripts by representational difference analysis. *Nucleic Acids Research*, **25**, 2681–2682.

ORTON, T. C., ADAM, H. K., BENTLEY, M., HOLLOWAY, B. and TUCKER, M. J., 1984, Clobuzarit: species differences in the morphological and biochemical response of the liver following chronic administration. *Toxicology and Applied Pharmacology*, **73**, 138–151.

PELKONEN, O., MAENPAA, J., TAAVITSAINEN, P., RAUTIO, A. and RAUNIO, H., 1998, Inhibition and Induction of human cytochrome P450 (CYP) enzymes. *Xenobiotica*, **28**, 1203–1253.

PHILIPS, S. M., BENDALL, A. J. and RAMSHAW, I. A., 1990, Isolation of genes associated with high metastatic potential in rat mammary adenocarcinomas. *Journal of the National Cancer Institute*, **82**, 199–203.

PRASHAR, Y. and WEISSMAN, S. M., 1996, Analysis of differential gene expression by display of 3'end restriction fragments of cDNAs. *Proceedings of the National Academy of Sciences (USA)*, **93**, 659–663.

RAGNO, S., ESTRADA, I., BUTLER, R. and COLSTON, M. J., 1997, Regulation of macrophage gene expression following invasion by *Mycobacterium tuberculosis*. *Immunology Letters*, **57**, 143–146.

RAMANA, K. V. and KOHLI, K. K., 1998, Gene regulation of cytochrome P450—an overview. *Indian Journal of Experimental Biology*, **36**, 437–446.

RICHARD, L., VELASCO, P. and DETMAR, M., 1998, A simple immunomagnetic protocol for the selective isolation and long-term culture of human dermal microvascular endothelial cells. *Experimental Cell Research*, **240**, 1–6.

ROCKETT, J. C., ESDAILE, D. J. and GIBSON, G. G., 1997, Molecular profiling of non-genotoxic hepatocarcinogenesis using differential display reverse transcription-polymerase chain reaction (ddRT-PCR). *European Journal of Drug. Metabolism and Pharmacokinetics*, **22**, 329–333.

RODRICKS, J. V. and TURNBULL, D., 1987, Inter-species differences in peroxisomes and peroxisome proliferation. *Toxicology and Industrial Health*, **3**, 197–212.

ROGLER, G., HAUSMANN, M., VOGL, D., ASCHENBRENNER, E., ANDUS, T., FALK, W., ANDREESEN, R., SCHOLMERICH, J. and GROSS, V., 1998, Isolation and phenotypic characterization of colonic macrophages. *Clinical and Experimental Immunology*, **112**, 205–215.

ROHN, W. M., LEE, Y. J. and BENVENISTE, E. N., 1996, Regulation of class II MHC expression. *Critical Reviews in Immunology*, **16**, 311–330.

RUDIN, C. M. and THOMPSON, C. B., 1998, B-cell development and maturation. *Seminars in Oncology*, **25**, 435–446.

SAKAGUCHI, N., BERGER, C. N. and MELCHERS, F., 1986, Isolation of a cDNA copy of an RNA species expressed in murine pre-B cells. *EMBO Journal*, **5**, 2139–2147.

SAMBROOK, J., FRITSCH, E. F. and MANIATIS, T., 1989, Gel electrophoresis of DNA. In N. Ford, M. Nolan and M. Fergusen (eds), *Molecular Cloning—A laboratory manual*, 2nd edition (New York: Cold Spring Harbour Laboratory Press), Volume 1, pp. 6–37.

SARGENT, T. D. and DAWID, I. B., 1983, Differential gene expression in the gastrula of Xenopus laevis. *Science*, **222**, 135–139.

SCHENA, M., SHALON, D., HELLER, R., CHAI, A., BROWN., P. O. and DAVIS, R. W., 1996, Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes. *Proceedings of the National Academy of Sciences (USA)*, **93**, 10614–10619.

SCHNEIDER, C., KING, R. M. and PHILIPSON, L., 1988, Genes specifically expressed at growth arrest of mammalian cells. *Cell*, **54**, 787–793.

SCHNEIDER-MAUNOURY, S., GILARDI-HEBENSTREIT, P. and CHARNAY, P., 1998, How to build a vertebrate hindbrain. Lessons from genetics. *C R Academy of Science III*, **321**, 819–834.

SEMENZA, G. L., 1994, Transcriptional regulation of gene expression: mechanisms and pathophysiology. *Human Mutations*, **3**, 180–199.

SEWALL, C. H., BELL, D. A., CLARK, G. C., TRITSCHER, A. M., TULLY, D. B., VANDEN HEUVEL, J. and LUCIER, G. W., 1995, Induced gene transcription: implications for biomarkers. *Clinical Chemistry*, **41**, 1829–1834.

SINGH, N., AGRAWAL, S. and RASTOGI, A. K., 1997, Infectious diseases and immunity: special reference to major histocompatibility complex. *Emerging Infectious Diseases*, **3**, 41–49.

SMITH, N. R., LI, A., ALDERSLEY, M., HIGH, A. S., MARKHAM, A. F. and ROBINSON, P. A., 1997, Rapid determination of the complexity of cDNA bands extracted from DDRT-PCR polyacrylamide gels. *Nucleic Acids* Research, **25**, 3552–3554.

SOMPAYRAC, L., JANE, S., BURN., T. C., TENEN, D. G. and DANNA, K. J., 1995, Overcoming limitations of the mRNA differential display technique. *Nucleic Acids Research*, **23**, 4738–4739.

ST JOHN, T. P. and DAVIS, R. W., 1979, Isolation of galactose-inducible DNA sequences from Saccharomyces cerevisiae by differential plaque filter hybridisation. *Cell*, **16**, 443–452.

SUN, Y., HEGAMYER, G. and COLBURN, N. H., 1994, Molecular cloning of five messenger RNAs differentially expressed in preneoplastic or neoplastic JB6 mouse epidermal cells: one is homologous to human tissue inhibitor of metalloproteinases-3. *Cancer Research*, **54**, 1139–1144.

SUNG, Y. J. and DENMAN, R. B., 1997, Use of two reverse transcriptases eliminates false-positive results in differential display. *Biotechniques*, **23**, 462–464.

SUTTON, G., WHITE, O., ADAMS, M. and KERLAVAGE, A., 1995, TIGR Assembler: A new tool for assembling large shotgun sequencing projects. *Genome Science and Technology*, **1**, 9–19.

SUZUKI, Y., SEKIYA, T. and HAYASHI, K., 1991, Allele-specific polymerase chain reaction: a method for amplification and sequence determination of a single component among a mixture of sequence variants. *Analytical Biochemistry*, **192**, 82–84.

SYED, V., GU, W. and HECHT, N. B., 1997, Sertoli cells in culture and mRNA differential display provide a sensitive early warning assay system to detect changes induced by xenobiotics. *Journal of Andrology*, **18**, 264–273.

UITTERLINDEN, A. G., SLAGBOOM, P., KNOOK, D. L. and VIJGL, J., 1989, Two-dimensional DNA fingerprinting of human individuals. *Proceedings of the National Academy of Sciences (USA)*, **86**, 2742–2746.

ULLMAN, K. S., NORTHROP, J. P., VERWEIJ, C. L. and CRABTREE, G. R., 1990, Transmission of signals from the T lymphocyte antigen receptor to the genes responsible for cell proliferation and immune function: the missing link. *Annual Review of Immunology*, **8**, 421–452.

VASMATZIS, G., ESSAND, M., BRINKMANN, U., LEE, B. and PASTON, I., 1998, Discovery of three genes specifically expressed in human prostate by expressed sequence tag database analysis. *Proceedings of the National Academy of Sciences (USA)*, **95**, 300–304.

VELCULESCU, V. E., ZHANG, L., VOGELSTEIN, B. and KINZLER, K. W., 1995, Serial analysis of gene expression. *Science*, **270**, 484–487.

VOELTZ, G. K. and STEITZ, J. A., 1998, AuuuA sequences direct mRNA deadenylation uncoupled from decay during Xenopus early development. *Molecular and Cell Biology*, **18**, 7537–7545.

VOGELSTEIN, B. and KINZLER, K. W., 1993, The multistep nature of cancer. *Trends in Genetics*, **9**, 138–141.

WALTER, J., BELFIELD, M., HAMPSON, I. and READ, C., 1997, A novel approach for generating subtractive probes for differential screening by CCLS. *Life Science News*, **21**, 13–14.

WAN, J. S., SHARP, S. J., POIRIER, G. M.-C., WAGAMAN, P. C., CHAMBERS, J., PYATI, J., HOM, Y.-L., GALINDO, J. E., HUVAR, A., PETERSON, P. A., JACKSON, M. R. and ERLANDER, M. G., 1996, Cloning differentially expressed mRNAs. *Nature Biotechnology*, **14**, 1685–1691.

WALTER, J., BELFIELD, M., HAMPSON, I. and READ, C., 1997, A novel approach for generating subtractive probes for differential screening by CCLS, *Life Science News*, **21**, 13–14.

WANG, Z. and BROWN, D. D. 1991, A gene expression screen. *Proceedings of the National Academy of Sciences (USA)*, **88**, 11505–11509.

WAWER, C., RUGGEBERG, H., MEYER, G. and MUYZER, G., 1995, A simple and rapid electrophoresis method to detect sequence variation in PCR-amplified DNA fragments. *Nucleic Acids Research*, **23**, 4928–4929.

WELSH, J., CHADA, K., DALAL, S. S., CHENG, R., RALPH, D. and MCCLELLAND, M., 1992, Arbitrarily primed PCR fingerprinting of RNA. *Nucleic Acids Research*, **20**, 4965–4970.

WONG, H., ANDERSON, W. D., CHENG, T. and RIABOWOL, K. T., 1994, Monitoring mRNA expression by polymerase chain reaction: the 'primer-dropping' method. *Analytical Biochemistry*, **223**, 251–258.

WONG, K. K. and MCCLELLAND, M., 1994, Stress-inducible gene of Salmonella typhimurium identified by arbitrarily primed PCR of RNA. *Proceedings of the National Academy of Sciences (USA)*, **91**, 639–643.

WYNFORD-THOMAS, D., 1991, Oncogenes and anti-oncogenes: the molecular basis of tumour behaviour. *Journal of Pathology*, **165**, 187–201.

XHU, D., CHAN, W. L., LEUNG, B. P., HUANG, F. P., WHEELER, R., PIEDRAFITA, D., ROBINSON, J. H. and LIEW, F. Y., 1998, Selective expression of a stable cell surface molecule on type 2 but not type 1 helper T cells. *Journal of Experimental Medicine*, **187**, 787–794.

YANG, M. and SYTOWSKI, A. J., 1996, Cloning differentially expressed genes by linker capture subtraction. *Analytical Biochemistry*, **237**, 109–114.

ZHAO, N., HASHIDA, H., TAKAHASHI, N., MISUMI, Y. and SAKAKI, Y., 1995, High-density cDNA filter analysis: a novel approach for large scale quantitative analysis of gene expression. *Gene*, **156**, 207–213.

ZHAO, X. J., NEWSOME, J. T. and CIHLAR, R. L., 1998, Up-regulation of two *candida albicans* genes in the rat model of oral candidiasis detected by differential display. *Microbial Pathogenesis*, **25**, 121–129.

ZIMMERMANN, C. R., ORR, W. C., LECLERC, R. F., BARNARD, C. and TIMBERLAKE, W. E., 1980, Molecular cloning and selection of genes regulated in *Aspergillus* development. *Cell*, **21**, 709–715.

Subject: RE: [Fwd: Toxicology Chip]
Date: Mon, 3 Jul 2000 08:09:45 -0400
From: "Afshari.Cynthia" <afshari@niehs.nih.gov>
To: "'Diana Hamlet-Cox'" <dianahc@incyte.com>

You can see the list of clones that we have on our tox chip at
http://manuel.niehs.nih.gov/maps/guest/clonesrch.cfm
We selected a subset of genes (2000K) that we believed critical to tox response and basic cellular processes and added a set of clones and ESTs to this. We have included a set of control genes (80+) that were selected by the NHGRI because they did not change across a large set of array experiments. However, we have found that some of these genes change significantly after tox treatments and are in the process of looking at the variation of each of these 80+ genes across our experiments.
Our chips are constantly changing and being updated and we hope that our data will lead us to what the toxchip should really be.
I hope this answers your question.
Cindy Afshari

> ----------
> *From:* Diana Hamlet-Cox
> *Sent:* Monday, June 26, 2000 8:52 PM
> *To:* afshari@niehs.nih.gov
> *Subject:* [Fwd: Toxicology Chip]
>
> Dear Dr. Afshari,
>
> Since I have not yet had a response from Bill Grigg, perhaps he was not the right person to contact.
>
> Can you help me in this matter? I don't need to know the sequences, necessarily, but I would like very much to know what types of sequences are being used, e.g., GPCRs (more specific?), ion channels, etc.
>
> Diana Hamlet-Cox
>
> -------- Original Message --------
> Subject: Toxicology Chip
> Date: Mon, 19 Jun 2000 18:31:48 -0700
> From: Diana Hamlet-Cox <dianahc@incyte.com>
> Organization: Incyte Pharmaceuticals
> To: grigg@niehs.nih.gov
>
> Dear Colleague:
>
> I am doing literature research on the use of expressed genes as pharmacotoxicology markers, and found the Press Release dated February 29, 2000 regarding the work of the NIEHS in this area. I would like to know if there is a resource I can access (or you could provide?) that would give me a list of the 12,000 genes that are on your Human ToxChip Microarray. In particular, I am interested in the criteria used to select sequences for the ToxChip, including any control sequences included in the microarray.
>
> Thank you for your assistance in this request.
>
> Diana Hamlet-Cox, Ph.D.
> Incyte Genomics, Inc.
>
> --
>
> ==========

> This email message is for the sole use of the intended recipient s and
> may contain confidential and privileged information subject to
> attorney-client privilege. Any unauthorized review, use, disclosure or
> distribution is prohibited. If you are not the intended recipient,
> please contact the sender by reply email and destroy all copies of the
> original message.
> ==========
>
>

ELSEVIER

Toxicology Letters 112–113 (2000) 467–471

Toxicology Letters

www.elsevier.com/locate/toxlet

Expression profiling in toxicology — potentials and limitations

Sandra Steiner *, N. Leigh Anderson

Large Scale Biology Corporation, 9620 Medical Center Drive, Rockville, MD 20850-3338, USA

Abstract

Recent progress in genomics and proteomics technologies has created a unique opportunity to significantly impact the pharmaceutical drug development processes. The perception that cells and whole organisms express specific inducible responses to stimuli such as drug treatment implies that unique expression patterns, molecular fingerprints, indicative of a drug's efficacy and potential toxicity are accessible. The integration into state-of-the-art toxicology of assays allowing one to profile treatment-related changes in gene expression patterns promises new insights into mechanisms of drug action and toxicity. The benefits will be improved lead selection, and optimized monitoring of drug efficacy and safety in pre-clinical and clinical studies based on biologically relevant tissue and surrogate markers.

Keywords: Proteomics; Genomics; Toxicology

1. Introduction

The majority of drugs act by binding to protein targets, most to known proteins representing enzymes, receptors and channels, resulting in effects such as enzyme inhibition and impairment of signal transduction. The treatment-induced perturbations provoke feedback reactions aiming to compensate for the stimulus, which almost always are associated with signals to the nucleus, resulting in altered gene expression. Such gene expression regulations account for both the pharmacological action and the toxicity of a drug and can be visualized by either global mRNA or global protein expression profiling. Hence, for each individual drug, a characteristic gene regulation pattern, its molecular fingerprint, exists which bears valuable information on its mode of action and its mechanism of toxicity.

Gene expression is a multistep process that results in an active protein (Fig. 1). There exist numerous regulation systems that exert control at and after the transcription and the translation step. Genomics, by definition, encompasses the quantitative analysis of transcripts at the mRNA level, while the aim of proteomics is to quantify gene expression further down-stream, creating a snapshot of gene regulation closer to ultimate cell function control.

* Corresponding author. Tel.: +1-301-4245989; fax: +1-301-7624892.
E-mail address: steiner@lsbc.com (S. Steiner)

0378-4274/00/$ - see front matter
PII: S0378-4274(99)00236-2

2. Global mRNA profiling

Expression data at the mRNA level can be produced using a set of different technologies such as DNA microarrays, reverse transcript imaging, amplified fragment length polymorphism (AFLP), serial analysis of gene expression (SAGE) and others. Currently, DNA microarrays are very popular and promise a great potential. On a typical array, each gene of interest is represented either by a long DNA fragment (200–2400 bp) typically generated by polymerase chain reaction (PCR) and spotted on a suitable substrate using robotics (Schena et al., 1995; Shalon et al., 1996) or by several short oligonucleotides (20–30 bp) synthesized directly onto a solid support using photolabile nucleotide chemistry (Fodor et al., 1991; Chee et al., 1996). From control and treated tissues, total RNA or mRNA is isolated and reverse transcribed in the presence of radioactive or fluorescent labeled nucleotides, and the labeled probes are then hybridized to the arrays. The intensity of the array signal is measured for each gene transcript by either autoradiography or laser scanning confocal microscopy. The ratio between the signals of control and treated samples reflect the relative drug-induced change in transcript abundance.

3. Global protein profiling

Global quantitative expression analysis at the protein level is currently restricted to the use of two-dimensional gel electrophoresis. This technique combines separation of tissue proteins by isoelectric focusing in the first dimension and by sodium dodecyl sulfate slab gel electrophoresis-based molecular weight separation on the second, orthogonal dimension (Anderson et al., 1991). The product is a rectangular pattern of protein spots that are typically revealed by Coomassie Blue, silver or fluorescent staining (Fig. 2). Protein spots are identified by mass spectrometry following generation of peptide mass fingerprints (Mann et al., 1993) and sequence tags (Wilkins et al., 1996). Similar to the mRNA approach, the ratio between the optical density of spots from control and treated samples are compared to search for treatment-related changes.

4. Expression data analysis

Bioinformatics forms a key element required to organize, analyze and store expression data from either source, the mRNA or the protein level. The overall objective, once a mass of high-quality

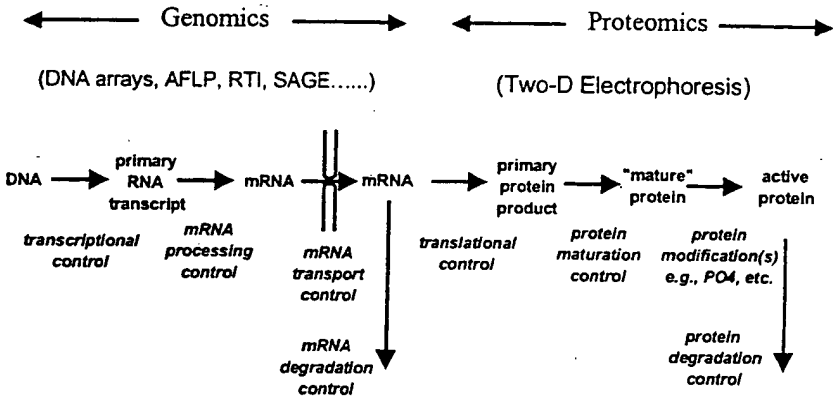


Fig. 1. Production of an active protein is a multistep process in which numerous regulation systems exert control at various stages of expression. Molecular fingerprints of drugs can be visualized through expression profiling at the mRNA level (genomics) using a variety of technologies and at the protein level (proteomics) using two-dimensional gel electrophoresis.

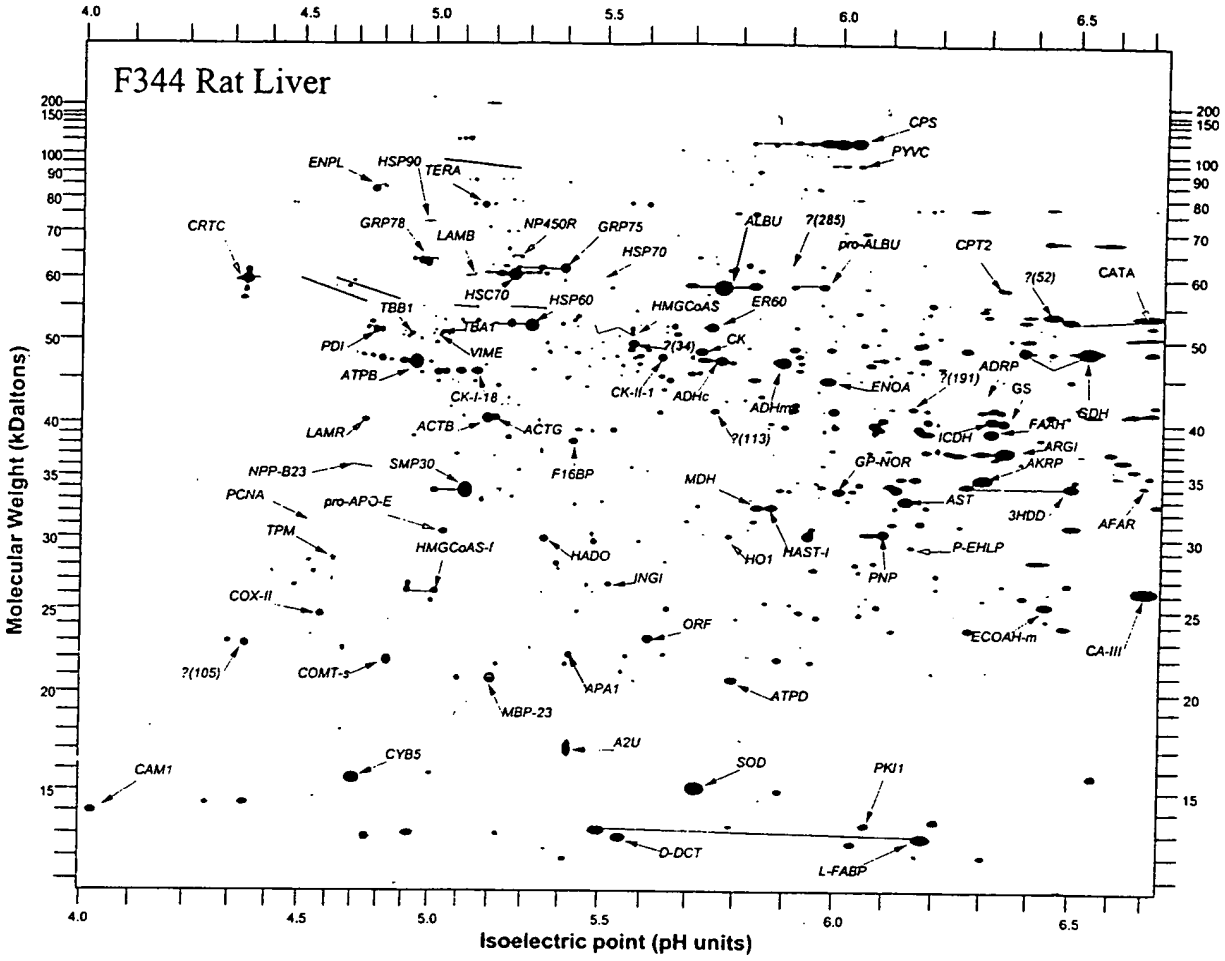


Fig. 2. Computerized representation of a Coomassie Blue stained two-dimensional gel electrophoresis pattern of Fischer F344 rat liver homogenate.

quantitative expression data has been collected, is to visualize complex patterns of gene expression changes, to detect pathways and sets of genes tightly correlated with treatment efficacy and toxicity, and to compare the effects of different sets of treatment (Anderson et al., 1996). As the drug effect database is growing, one may detect similarities and differences between the molecular fingerprints produced by various drugs, information that may be crucial to make a decision whether to refocus or extend the therapeutic spectrum of a drug candidate.

5. Comparison of global mRNA and protein expression profiling

There are several synergies and overlaps of data obtained by mRNA and protein expression analysis. Low abundant transcripts may not be easily quantified at the protein level using standard two-dimensional gel electrophoresis analysis and their detection may require prefractionation of samples. The expression of such genes may be preferably quantified at the mRNA level using techniques allowing PCR-mediated target amplifi-

cation. Tissue biopsy samples typically yield good quality of both mRNA and proteins; however, the quality of mRNA isolated from body fluids is often poor due to the faster degradation of mRNA when compared with proteins. RNA samples from body fluids such as serum or urine are often not very 'meaningful', and secreted proteins are likely more reliable surrogate markers for treatment efficacy and safety. Detection of post-translational modifications, events often related to function or nonfunction of a protein, is restricted to protein expression analysis and rarely can be predicted by mRNA profiling. Information on subcellular localization and translocation of proteins has to be acquired at the level of the protein in combination with sample prefractionation procedures. The growing evidence of a poor correlation between mRNA and protein abundance (Anderson and Seilhamer, 1997) further suggests that the two approaches, mRNA and protein profiling, are complementary and should be applied in parallel.

6. Expression profiling and drug development

Understanding the mechanisms of action and toxicity, and being able to monitor treatment efficacy and safety during trials is crucial for the successful development of a drug. Mechanistic insights are essential for the interpretation of drug effects and enhance the chances of recognizing potential species specificities contributing to an improved risk profile in humans (Richardson et al., 1993; Steiner et al., 1996b; Aicher et al., 1998). The value of expression profiling further increases when links between treatment-induced expression profiles and specific pharmacological and toxic endpoints are established (Anderson et al., 1991, 1995, 1996; Steiner et al. 1996a). Changes in gene expression are known to precede the manifestation of morphological alterations, giving expression profiling a great potential for early compound screening, enabling one to select drug candidates with wide therapeutic windows reflected by molecular fingerprints indicative of high pharmacological potency and low toxicity (Arce et al., 1998). In later phases of drug development, surrogate markers of treatment efficacy and toxicity can be applied to optimize the monitoring of pre-clinical and clinical studies (Doherty et al., 1998).

7. Perspectives

The basic methodology of safety evaluation has changed little during the past decades. Toxicity in laboratory animals has been evaluated primarily by using hematological, clinical chemistry and histological parameters as indicators of organ damage. The rapid progress in genomics and proteomics technologies creates a unique opportunity to dramatically improve the predictive power of safety assessment and to accelerate the drug development process. Application of gene and protein expression profiling promises to improve lead selection, resulting in the development of drug candidates with higher efficacy and lower toxicity. The identification of biologically relevant surrogate markers correlated with treatment efficacy and safety bears a great potential to optimize the monitoring of pre-clinical and clinical trails.

References

Aicher, L., Wahl, D., Arce, A., Grenet, O., Steiner, S., 1998. New insights into cyclosporine A nephrotoxicity by proteome analysis. Electrophoresis 19, 1998–2003.

Anderson, N.L., Seilhamer, J., 1997. A comparison of selected mRNA and protein abundances in human liver. Electrophoresis 18, 533–537.

Anderson, N.L., Esquer-Blasco, R., Hofmann, J.P., Anderson, N.G., 1991. A two-dimensional gel database of rat liver proteins useful in gene regulation and drug effects studies. Electrophoresis 12, 907–930.

Anderson, L., Steele, V.K., Kelloff, G.J., Sharma, S., 1995. Effects of oltipraz and related chemoprevention compounds on gene expression in rat liver. J. Cell. Biochem. Suppl. 22, 108–116.

Anderson, N.L., Esquer-Blasco, R., Richardson, F., Foxworthy, P., Eacho, P., 1996. The effects of peroxisome proliferators on protein abundances in mouse liver. Toxicol. Appl. Pharmacol. 137, 75–89.

Arce, A., Aicher, L., Wahl, D., Esquer-Blasco, R., Anderson, N.L., Cordier, A., Steiner, S., 1998. Changes in the liver proteome of female Wistar rats treated with the hypoglycemic agent SDZ PGU 693. Life Sci. 63, 2243–2250.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X.C., Stern, D., Winkler, J., Lockhart, D.J., Morris, M.S., Fodor, S.P., 1996. Accessing genetic information with high-density DNA arrays. Science 274, 610–614.

Doherty, N.S., Littman, B.H., Reilly, K., Swindell, A.C., Buss, J., Anderson, N.L., 1998. Analysis of changes in acute-phase plasma proteins in an acute inflammatory response and in rheumatoid arthritis using two-dimensional gel electrophoresis. Electrophoresis 19, 355–363.

Fodor, S.P., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T., Solas, D., 1991. Light-directed, spatially addressable parallel chemical synthesis. Science 251, 767–773.

Mann, M., Hojrup, P., Roepsdorff, P., 1993. Use of mass spectrometric molecular weight information to identify proteins in sequence databases. Biol. Mass Spectrom. 22, 338–345.

Richardson, F.C., Strom, S.C., Copple, D.M., Bendele, R.A., Probst, G.S., Anderson, N.L., 1993. Comparisons of protein changes in human and rodent hepatocytes induced by the rat-specific carcinogen, methapyrilene. Electrophoresis 14, 157–161.

Schena, M., Shalon, D., Davis, R.W., Brown, P.O., 1995. Quantitative monitoring of gene expresssion patterns with a complementary DNA microarray. Science 251, 467–470.

Shalon, D., Smith, S.J., Brown, P.O., 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res. 6, 639–645.

Steiner, S., Wahl, D., Mangold, B.L.K., Robison, R., Rayrnackers, J., Meheus, L., Anderson, N.L., Cordier, A., 1996a. Induction of the adipose differentiation-related protein in liver of etomoxir treated rats. Biochem. Biophys. Res. Commun. 218, 777–782.

Steiner, S., Aicher, L., Raymackers, J., Meheus, L., Esquer-Blasco, R., Anderson, L., Cordier, A., 1996b. Cyclosporine A mediated decrease in the rat renal calcium binding protein calbindin-D 28 kDa. Biochem. Pharmacol. 51, 253–258.

Wilkins, M.R., Gasteiger, E., Sanchez, J.C., Appel, R.D., Hochstrasser, D.F., 1996. Protein identification with sequence tags. Curr. Biol. 6, 1543–1544.

A

Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 10614–10619, October 1996
Biochemistry

Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes

(Human Genome Project/DNA chip/gene discovery/T cell)

MARK SCHENA*†, DARI SHALON‡, RENU HELLER*, ANDREW CHAI*, PATRICK O. BROWN§, AND RONALD W. DAVIS*

*Department of Biochemistry, Beckman Center, Stanford University Medical Center, Stanford, CA 94305; ‡Synteni, Palo Alto, CA 94306; and §Department of Biochemistry and Howard Hughes Medical Institute, Beckman Center, Stanford University Medical Center, Stanford, CA 94305

Contributed by Ronald W. Davis, June 26, 1996

ABSTRACT **Microarrays containing 1046 human cDNAs of unknown sequence were printed on glass with high-speed robotics. These 1.0-cm² DNA "chips" were used to quantitatively monitor differential expression of the cognate human genes using a highly sensitive two-color hybridization assay. Array elements that displayed differential expression patterns under given experimental conditions were characterized by sequencing. The identification of known and novel heat shock and phorbol ester-regulated genes in human T cells demonstrates the sensitivity of the assay. Parallel gene analysis with microarrays provides a rapid and efficient method for large-scale human gene discovery.**

Biology has entered the genome era (1). Complete genome sequences for all of the model organisms and human will probably be available by the year 2003 (2). Torrents of human expressed sequence tags (ESTs) provide a starting point for elucidating the function of tens of thousands of cognate genes (3). Genome analysis will provide insights into growth, development, differentiation, homeostasis, aging, and the onset of diseases (1–3). A detailed understanding of the human genome will require the implementation of sophisticated methods for gene expression analysis and gene discovery.

Recently, a microarray-based method for high-throughput monitoring of plant gene expression was described (4). This "chip"-based approach involved using microarrays of cDNA clones as gene-specific hybridization targets to quantitatively measure expression of the corresponding plant genes (4, 5). A two-color fluorescence labeling and detection scheme facilitated sensitive differential expression analysis of different plant tissues (4, 5). The efficiency of this approach for studies in higher plants suggested the use of this method for human genome analysis (4–7). Here, we report the use of cDNA microarrays for human gene expression monitoring, biological investigation, and gene discovery.

MATERIALS AND METHODS

Human cDNA Clones. The cDNA library was made with mRNA from human peripheral blood lymphocytes transformed with the Epstein–Barr virus. Inserts >600 bp were cloned into the lambda vector λYES-R to generate 10^7–10^8 recombinants. Bacterial transformants were obtained by infecting *E. coli* strain JM107/λKC. Colonies were picked at random and propagated in a 96-well format, and minilysate DNA was prepared by alkaline lysis using REAL preps (Qiagen, Chatsworth, CA). Inserts were amplified by PCR in a 96-well format using primers (PAN132, 5'-CCTCTATACTTTAACGTCAAGG; and PAN133, 5'-TTGTGTGGAATTGTGAGCGG) complementary to the λYES polylinker and containing a six-carbon amino modification (Glen Research, Sterling, VA) on the 5' end. PCR products were purified in a 96-well format using QIAquick columns (Qiagen).

Microarray Preparation. Amino-modified PCR products were suspended at a concentration of 0.5 mg/ml in 3× standard saline citrate (SSC) and arrayed from 96-well microtiter plates onto silylated microscope slides (CEL Associates, Houston) using high-speed robotics (4–7). A total of 1056 cDNAs, representing 1046 human clones and 10 *Arabidopsis* controls, were arrayed in 1.0-cm² areas. Printed arrays were incubated for 4 hr in a humid chamber to allow rehydration of the array elements and rinsed, once in 0.2% SDS for 1 min, twice in H_2O for 1 min, and once for 5 min in sodium borohydride solution (1.0 g of $NaBH_4$ dissolved in 300 ml of PBS and 100 ml of 100% ethanol). The arrays were submerged in H_2O for 2 min at 95°C, transferred quickly into 0.2% SDS for 1 min, rinsed twice in H_2O, air dried, and stored in the dark at 25°C.

Fluorescent Probes. Tissue mRNAs were purchased (CLONTECH). Jurkat mRNA was isolated as described by Schena *et al.* (4). Probes were made as described (4) with several modifications. The reverse transcriptase used here was Superscript II RNase H− (GIBCO). The Cy5-dCTP was purchased from Amersham. Each reverse transcription reaction contained 3.0 μg of total human mRNA. *Arabidopsis* control mRNAs were made by *in vitro* transcription of cloned HAT4, HAT22, and YesAt-23 cDNAs (4, 8, 9) using an RNA Transcription Kit (Stratagene). For quantitation, the mRNAs were doped into the reverse transcription reaction at ratios of 1:100,000, 1:10,000, and 1:1000 (wt/wt) respectively. Following the reverse transcription step, samples were treated with 2.5 μl of 1 M sodium hydroxide for 10 min at 37°C, then neutralized by adding 2.5 μl of 1 M Tris·HCl (pH 6.8) and 2.0 μl of 1 M HCl. Probe mixtures contained cDNA products derived from 3 μg of total mRNA, suspended in 5.0 μl of hybridization buffer (5× SSC plus 0.2% SDS).

Hybridization and Scanning. Probes were hybridized to 1.0-cm² microarrays under a 14 × 14 mm glass coverslip for 6–12 hr at 60°C in a custom-built hybridization chamber (4–7). Arrays were washed for 5 min at room temperature (25°C) in low stringency wash buffer (1× SSC/0.2% SDS), then for 10 min at room temperature in high stringency wash buffer (0.1× SSC/0.2% SDS). Arrays were scanned in 0.1× SSC using a fluorescence laser scanning device (4–7), fitted with a custom filter set (Chroma Technology, Brattleboro, VT). Accurate differential expression measurements (i.e., final fluorescence ratios) were obtained by taking the average of the ratios of two independent hybridizations.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: EST, expressed sequence tag.

Data deposition: The sequences reported in this paper have been deposited in the GenBank data base (accession nos. U56654–U56660).

†To whom reprint requests should be addressed. e-mail: schena@cmgm.stanford.edu.

Cell Culture. Jurkat cells were grown in a tissue culture incubator (37°C and 5% CO_2) in RPMI medium supplemented with 10% fetal bovine serum, 100 μg of streptomycin per ml, and 500 units of penicillin per ml. Heat shock corresponded to a 4-hr incubation at 43°C. Phorbol ester treated cells were grown for 4 hr in the presence of 50 ng of phorbol 12-myristate 13-acetate (PMA) per ml.

RNA Blotting. Dot blots were performed as described (4).

DNA Sequencing. Sequences were obtained using the PAN132 and PAN133 primers and a 373A automated sequencer, according to the instructions of the manufacturer (Applied Biosystems).

Computer Graphics and Informatics. Pseudocolor representations of fluorescent images were made with National Institutes of Health IMAGE software (version 1.52). Software for differential expression representations was purchased from Imaging Research (St. Catherine's, ON, Canada). Sequence searches were made to the nonredundant nucleotide data base at the National Center for Biotechnology Information (NCBI) using Macintosh BLAST software. The EST data base was accessed via the World Wide Web (http://www.ncbi.nlm.nih.gov/).

RESULTS

Gene Discovery and the Heat Shock Response. Microarrays were used to examine the heat shock response in cultured human T (Jurkat) cells. Control (37°C) and heat-treated (43°C) cells were harvested and lysed, and total mRNA from the two cell samples was labeled by reverse transcriptase incorporation of fluorescein- and Cy5-dCTP, respectively. In a second set of labeling reactions, the fluorescent groups were "swapped" such that samples from control and heat-treated samples were labeled with Cy5- and fluorescein-dCTP, respectively. Each pair of fluorescent probes was hybridized to a 1056-element microarray. The arrays were washed at high stringency and scanned with a confocal laser scanning device to detect emission of the two fluorescent groups.

Hybridization signals were observed to >95% of the human cDNA array elements, but not to any of the *Arabidopsis* negative controls (Fig. 1). Fluorescence intensities spanned more than three orders of magnitude for the 1046 array elements surveyed (Fig. 1). Comparative expression analysis of heat shocked versus control cells in the two experiments revealed 17 array elements that displayed altered fluorescence ratios of ≥2.0-fold (Figs. 1 and 2-4). Of the 17 putative differentially expressed genes, 11 were induced by heat shock treatment and 6 displayed modest repression (Figs. 1 and 2-4).

To determine the identity of the heat-regulated genes, cDNAs corresponding to each of the 17 array elements were sequenced on the proximal and distal end. Data base searches revealed perfect matches for 14 of the 17 clones, and in each case proximal and distal cDNA sequences mapped to the same gene (Table 1). Of the 1046 human genes examined on the microarray, the five most highly induced in heat-treated cells were heat shock protein 90α (hsp90α), dnaJ, hsp90β, polyubiquitin, and t-complex polypeptide-1 (tcp-1) (Table 1). Three of the 17 clones did not match any entry in the public data base, though one of the clones (B7) exhibited significant homology to an EST from *Caenorhabditis elegans* (Table 1). Each of the novel sequences (B7-B9) exhibited ~2-fold induction (Table 1) and relatively low-level expression (Table 2).

To confirm the microarray results, mRNA levels for each of the genes were measured by RNA blotting. Each of the genes that displayed heat shock induction, including the three novel

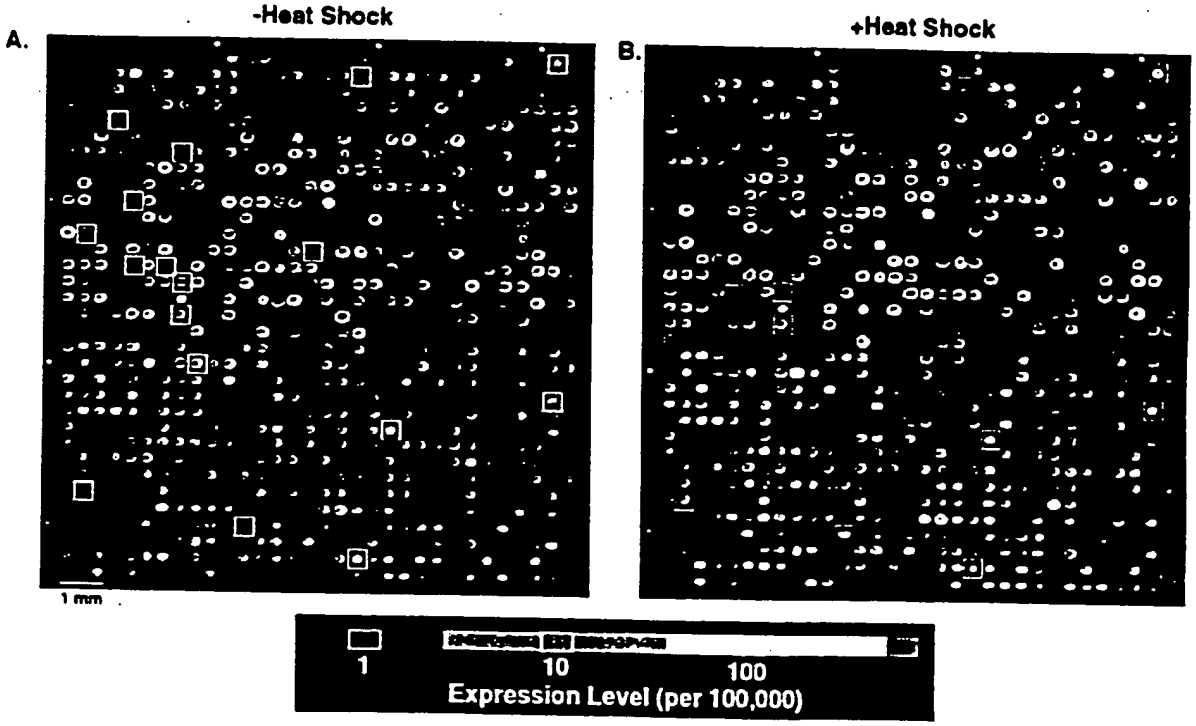


FIG. 1. Human gene expression monitored on a microarray. Fluorescent scans represented in a pseudocolor scale correspond to expression levels. The array contains 10 *Arabidopsis* controls (upper left corner, elements 1–10) and 1046 human peripheral blood cDNAs. Fluorescent probes were prepared by labeling mRNA from Jurkat cells grown at 37°C (−Heat Shock, *A*) or 43°C (+Heat Shock, *B*). Array elements that display altered fluorescence intensity (white boxes) corresponded to genes activated (red boxes) or repressed (green boxes) by heat shock. The color bar was calibrated in separate experiments using known quantities (wt/wt) of *Arabidopsis* control mRNAs added to the labeling reaction. Microarray rows (at left) and columns (at the top) are demarcated at 10 element increments (white circles). (Bar = 1 mm.)

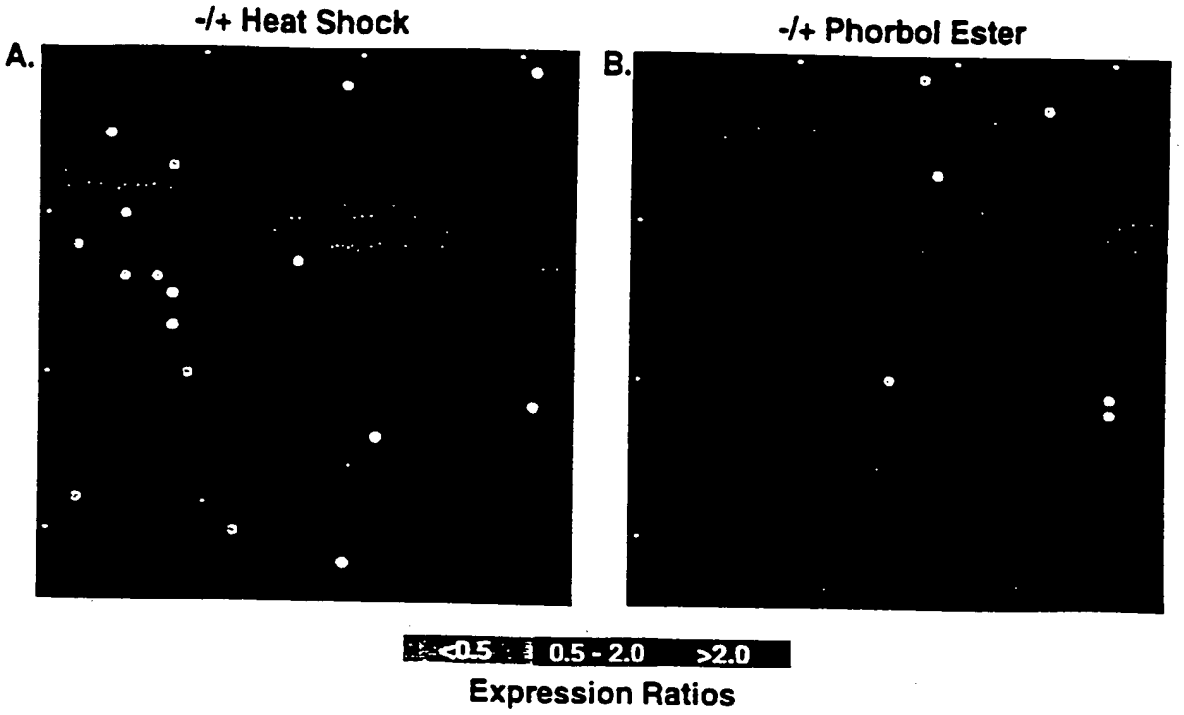


FIG. 2. Elemental displays of activated and repressed genes. Fluorescence ratios of two-color microarray scans (Fig. 1) are depicted schematically. Fluorescein-labeled probes from Jurkat cells subjected to (*A*) heat shock or (*B*) phorbol ester treatment were compared with Cy5-labeled probes from untreated cells. In a second set of reactions, the fluorescent groups were swapped (see text). The data represent the average of the ratios from two hybridizations, excluding values in which the difference of the two ratios was greater than half the average ratio. The color bar corresponds to expression ratios, which are independent of the absolute expression level of a given gene.

Table 1. Microarray elements corresponding to differentially expressed genes

Clone	Row	Column	Ratio	Blast identity	Accession no.
B1	24	21	0.5	CYC oxidase III	J01415, J01415
B2	1	31	0.5	β-Actin	NR, X00351
B3	15	8	0.5	CYC oxidase III	J01415, J01415
B4	32	19	0.5	CYC oxidase III	J01415, J01415
B5	17	8	0.5	CYC oxidase III	J01415, J01415
B6	22	31	0.5	β-Actin	NR, X00351
B7*	5	4	2.0	Novel†	U56653, U56654
B8	2	19	2.0	Novel†	U56655, U56656
B9	14	5	2.2	Novel†	U56657, U56658
B10	7	8	2.4	Polyubiquitin	X04803, X04803
B11	12	2	2.4	TCP-1	X52882, X52882
B12	28	2	2.5	Polyubiquitin	M17597, M17597
B13	14	7	2.5	Polyubiquitin	X04803, X04803
B14	20	9	2.6	HSP90β	M16660, M16660
B15	30	12	4.0	DnaJ homolog	D13388, D13388
B16	10	5	5.8	HSP90α	X07270, X07270
B17	13	16	6.3	HSP90α	M27024, X15183
B18	7	19	2.0	β2-microglobulin	S54761, M30683
B19	21	30	2.1	Novel†	U56659, U56660
B20	3	26	2.2	β2-microglobulin	S54761, M30683
B21	1	18	2.6	PGK	M11968, L00160
B22	22	30	3.5	NF-κB1	Z47744, M55643
B23	20	16	19	PAC-1	L11329, L11329

Clone name, array position (Fig. 1), fluorescence ratio, sequence identity, and acession number of cDNAs that manifested a differential expression pattern with probes prepared from heat shock- (B1-17) or phorbol ester-treated (B18-23) Jurkat cells. Clones showing >98% identity over 300 nucleotides were assumed to be identical to known sequences. All genes are nuclear except CYC oxidase III (mitochondrial). Accession numbers reflect the highest score for proximal and distal sequence traces, respectively. CYC, cytochrome *c*; TCP-1, T-complex polypeptide; HSP, heat shock protein; PGK, phosphoglycerate kinase; NF-κB, nuclear factor-kappaB; PAC-1, phosphatase of activated cells; and NR, trace not readable due to the presence of poly(A)+ tract.

*B7 is 67% identical to an EST from *C. elegans* (D76026).

†No match in the public data bases.

Table 2. Human gene expression monitored by microarray and RNA blot analyses

Clone	Blast identity	Expression level, per 10^5 mRNAs			
		Microarray	Ratio	RNA blot	Ratio
B1	CYC oxidase III	92/46	0.5	100/80	0.8
B2	β-Actin	240/120	0.5	270/280	1.0
B3	CYC oxidase III	36/18	0.5	ND	ND
B4	CYC oxidase III	76/38	0.5	ND	ND
B5	CYC oxidase III	62/31	0.5	ND	ND
B6	β-Actin	180/89	0.5	ND	ND
B7	Novel (weakly to D76026)	1.3/2.6	2.0	0.77/1.8	2.3
B8	Novel	2.0/4.0	2.0	1.5/3.4	2.3
B9	Novel	0.8/1.8	2.2	1.2/1.8	1.5
B10	Polyubiquitin	0.8/1.9	2.4	25/89	3.6
B11	TCP-1	2.3/5.5	2.4	7.1/27	3.8
B12	Polyubiquitin	0.8/2.0	2.5	ND	ND
B13	Polyubiquitin	1.7/4.3	2.5	ND	ND
B14	HSP90β	75/200	2.6	30/120	4.0
B15	DnaJ homolog	1.0/4.0	4.0	1.6/13	8.1
B16	HSP90α	0.6/3.5	5.8	3.2/29	9.1
B17	HSP90α	0.8/5.0	6.3	8.6/62	7.2
B18	β_2-microglobulin	1.0/2.0	2.0	5.4/15	2.8
B19	Novel	1.2/2.5	2.1	4.5/9.5	2.5
B20	β_2-microglobulin	2.7/5.9	2.2	ND	ND
B21	Phosphoglycerate kinase	2.4/6.2	2.6	4.7/9.2	2.0
B22	NF-KB1	1.7/6.0	3.5	0.65/4.7	7.2
B23	PAC-1	0.5/9.5	19	0.21/15	71

Shown are expression levels per 100,000 mRNAs (wt/wt) of genes assayed with a microarray (Fig. 1) or RNA blot. Ratios correspond to values from cells subjected to heat shock (B1-17) or phorbol ester treatment (B18-23) relative to untreated cells. Clone and gene names are given in Table 1. ND, not determined.

sequences, exhibited elevated mRNA levels by dot blot analysis (Table 2). In all cases, expression ratios as determined by the two procedures differed by <2-fold for the genes identified in the heat shock experiments (Table 2). The two assays differed more widely in terms of assessing absolute expression levels; nonetheless, absolute expression as monitored on a microarray typically correlated with RNA blots to within a factor of five (Table 2).

Phorbol Ester Signaling. To explore a signaling pathway distinct from the heat shock response, microarrays were used to examine the cellular effects of phorbol ester treatment. Jurkat cells were treated with phorbol ester, harvested, lysed, and used as a source of mRNA. Samples of mRNA from untreated or phorbol ester-stimulated cells were labeled with reverse transcriptase. The probes were mixed, hybridized to microarrays, and scanned for fluorescence emission of the two fluorescent groups. A total of six array elements displayed ≥2.0-fold elevated signals with probes from phorbol ester-treated cells relative to control samples (Fig. 2*B*).

To determine the identity of the phorbol ester-induced genes, clones corresponding to the six array elements were sequenced. Data base searches revealed perfect matches for five of the six sequences (Table 1). The two most highly induced genes were the *PAC-1* tyrosine phosphatase and nuclear factor-kappa B1 (*NF-κB1*); modest activation was observed for phosphoglycerate kinase and β_2-microglobulin (Table 1). One remaining clone (B19) did not match any entry in the public data base (Table 1). B19 displayed a 2.1-fold induction and, similar to the novel heat shock genes, a relatively low absolute expression level (Tables 1 and 2). All six of the phorbol ester-inducible genes displayed increased steady-state mRNA levels by RNA blotting (Table 2). *PAC-1* expression (Fig. 1; Table 2) defined a detection limit of ~1:500,000 for the assay.

Transcript Imaging in Human Tissues. To determine whether microarrays could be used to monitor expression in human tissues, probes were prepared from human bone marrow, brain, prostate, and heart by labeling each mRNA sample with Cy5-dCTP. In a separate reaction, a control probe was prepared by labeling Jurkat mRNA with fluorescein-dCTP. The four Cy5-labeled probes were each mixed with an aliquot of the fluorescein-labeled control sample, and the four mixtures were hybridized to separate microarrays. The arrays were washed and scanned for fluorescence emission, and hybridization signals for each of the tissues samples were normalized to the Jurkat control to generate an expression profile for each of the 1046 clones present on the array.

Detectable expression was observed for all 15 of the heat shock and phorbol ester-regulated genes in the four tissue types examined (Fig. 3). In general, the expression level of each gene in Jurkat cells correlated rather closely with expression in the four tissues (Table 2; Fig. 3). Genes encoding β-actin and cytochrome *c* oxidase, the two most highly expressed of the 15 genes in Jurkat cells (Table 2), were highly expressed in bone marrow, brain, prostate, and heart (Fig. 3*A*). Expression of cytochrome *c* oxidase, hsp90α, and the novel B7 sequence was significantly greater in heart than in the other tissues (Fig. 3).

DISCUSSION

Many of the heat shock genes identified in this study encode factors that function either as molecular "chaperones" (HSP90α, HSP90β, DnaJ, TCP-1) or as mediators of protein degradation (polyubiquitin). The identification of these sequences is consistent with the biochemical basis of heat shock induction (10–15). Proteins undergo denaturation at elevated temperatures, and those that fail to maintain proper conformation must be selectively degraded (10–15). It will be interesting to determine whether the three novel heat shock-inducible sequences (B7–B9) mediate protein folding and turnover or possess some other biochemical activity. Complete nucleotide sequence determination, conceptual translation, expression monitoring, and biochemical analysis should provide a detailed functional understanding of these genes.

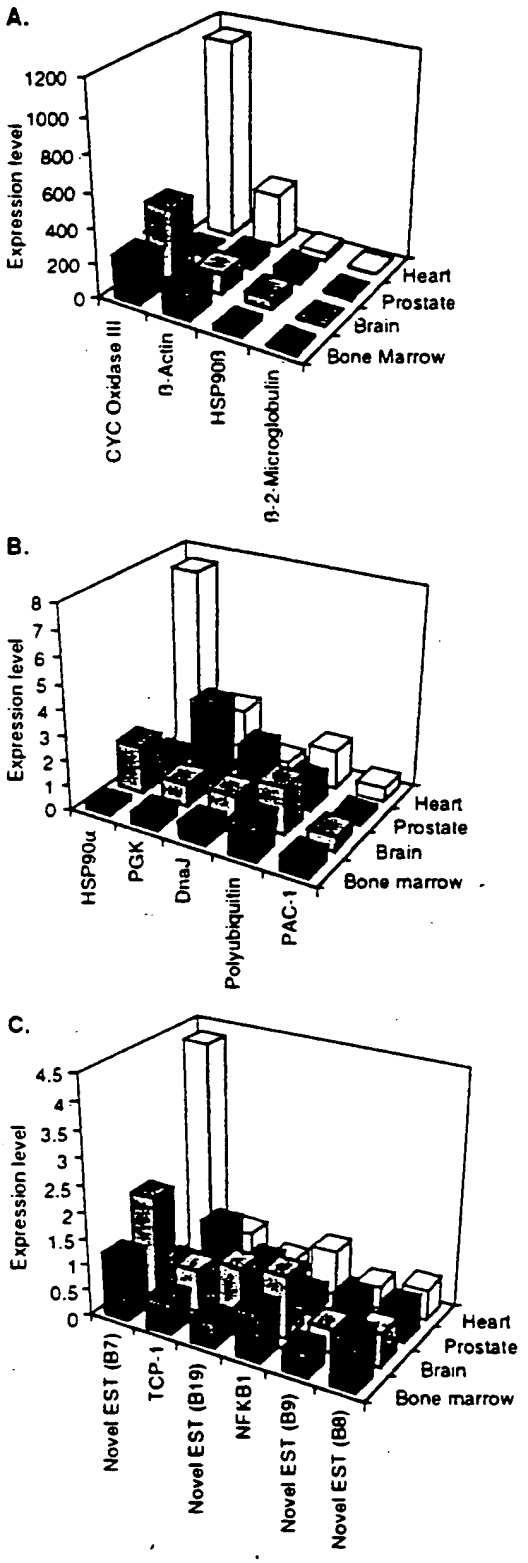


FIG. 3. Transcript profiles of heat shock and phorbol ester-regulated genes. Gene expression levels per 100,000 mRNAs (*x*-axes) are shown for 15 genes (Table 1) in human bone marrow (red), brain (green), prostate (blue), and heart (yellow). Genes are grouped according to expression levels (*A–C*).

Phorbol ester, a potent activator of protein kinase C (16, 17), induced a set of genes distinct from those involved in the heat shock pathway. The most highly induced gene identified in this study, *PAC-1*, encodes a nuclear tyrosine kinase that may play a role in regulating transcription and cell cycle progression (18). NF-κB1, a second phorbol ester-inducible gene, is an intensively studied member of the Rel transcription factor family (19–21). The Rel proteins are activated by a large number of stimuli, including phorbol esters, cytokines, bacterial and viral pathogens, and ultraviolet light (19–21). Modest activation was observed for three sequences not known to be inducible by phorbol esters, including phosphoglycerate kinase, β_2-microglobulin, and a novel human gene (B19). Extensive expression monitoring with microarrays should assist in understanding how each of these genes integrate into the highly complex phorbol ester signaling pathway.

It is striking that four novel human genes were discovered with an array of 1000 randomly chosen clones, particularly because the heat shock and phorbol ester signaling pathways have been so intensively studied (10–21). The facile discovery of these sequences underscores the fact that microarrays can be used for gene discovery in the absence of any sequence information. By this approach, clones are chosen at random from any library of interest and only those clones that display interesting expression patterns are sequenced and characterized. This parallel assay, coupled with a modest DNA sequencing facility, allows high-throughput human genome expression analysis and gene discovery.

Genes that are activated or repressed by a given stimulus provide functional clues to the cellular pathway involved (22–24). Detailed examination of these gene expression "signatures" can provide a dynamic view of the mode of action of a given signaling substance (22–24). Microarrays may thus allow rapid mechanistic examination of hormones, drugs, elicitors, and other small molecules; moreover, functional analysis of transcription factors, kinases, growth factors, cytokines, receptors, and other gene products should be possible. Efforts are underway to develop mRNA amplification strategies to enable probe preparation from minute tissue samples. This capability might allow for high-throughput patient screening in a clinical setting.

The current detection limit of the assay allows monitoring of transcripts that represent ~1:500,000 (wt/wt) of the total mRNA. This 10-fold increase in sensitivity compared with the original report (4) was achieved largely by modifying the coupling chemistry, which reduced background fluorescence. The significance of this improvement is considerable in that approximately half the human genes identified in this study, including all four novel sequences, exhibited expression levels below the original detection limit of 1:50,000 (4).

The ability to detect 2-fold changes in expression was achieved by the use of two-color fluorescence in the labeling and detection schemes, digitized data collection, and custom software. The importance of this capability is underscored by the fact that nearly all of the genes examined here exhibited <6-fold changes in expression. The four novel genes, which showed ≤2.2-fold activation, were probably overlooked in previous screens that used conventional differential expression techniques. It may be possible to further improve the precision of the microarray assay by the use of closely related fluorescent analogs, such as Cy3 and Cy5, in the labeling and hybridization reactions.

Microarrays offer a number of advantages over other potential high-capacity approaches to expression analysis. The chip-based approach enables small hybridization volumes, high array densities, and the use of fluorescence labeling and detection schemes. These features provide a set of performance specifications that are unattainable with filter-based approaches (25, 26). The use of cDNA clones provides hybridization specificity that is not readily attained with oligo-

nucleotide arrays (27–30). The parallel format of the assay provides a simultaneous differential expression readout for >1000 genes. This contrasts with sequencing-based methods, which require serial data collection for expression analysis (31, 32). A commercial source of cDNA microarrays would greatly speed the use of a chip-based approach to expression analysis.

The availability of large numbers of ESTs (3) provides a rich resource of human cDNA clones for microarraying. The >400,000 ESTs in the public data bases represent a significant subset of all human genes (3, 33). Microarrays of thousands of ESTs will provide a powerful analytical tool for future human gene expression studies. The ~100,000 genes in the human genome (2, 33) emphasize the need for microarrays of greater density. Attempts to improve microdeposition techniques are underway and should allow construction of arrays containing a complete set of human gene targets (http://cmgm.stanford.edu/~schena/). Microarrays of ~100,000 cDNA elements would allow expression monitoring of the entire human genome in a single hybridization. This capacity, coupled with detailed biochemical analysis of the individual gene products, would greatly speed the functional analysis of the human genome.

We thank S. Elledge (selledge@bcm.tmc.edu) for the human cDNA library, Qiagen representatives for help with plasmid purification, and A. J. Smith and colleagues at the Protein and Nucleic Acid (PAN) facility (Stanford) for oligonucleotide synthesis and DNA sequencing. We also thank members of the Davis, Brown, and Smith laboratories for critical comments and helpful discussions and Synteni employees for technical assistance. Support for R.W.D. was provided by the National Science Foundation (MCB9106011) and National Institutes of Health (R37HG00198) and for P.O.B. by the National Institutes of Health (3R21HG00450) and Howard Hughes Medical Institute. P.O.B. is an assistant investigator of the Howard Hughes Medical Institute.

1. Watson, J. D. (1993) *Gene* 135, 309–315.
2. Collins, F. S. (1995) *Proc. Natl. Acad. Sci. USA* 92, 10821–10823.
3. Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie, W. R. & Venter, J. C. (1991) *Science* 252, 1651–1656.
4. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. (1995) *Science* 270, 467–470.
5. Shalon, D. (1996) Ph.D. thesis (Stanford University).
6. Schena, M. (1996) *BioEssays* 18, 427–431.
7. Shalon, D., Smith, S. J. & Brown, P. O. (1996) *Genome Res.* 6, 639–645.
8. Schena, M. & Davis, R. W. (1994) *Proc. Natl. Acad. Sci. USA* 91, 8393–8397.
9. Schena, M. & Davis, R. W. (1992) *Proc. Natl. Acad. Sci. USA* 89, 3894–3898.
10. Jindal, S. (1996) *Trends Biotechnol.* 14, 17–20.
11. Wilkinson, K. D. (1995) *Annu. Rev. Nutr.* 15, 161–189.
12. Jakob, U. & Buchner, J. (1994) *Trends Biochem. Sci.* 19, 205–211.
13. Becker, J. & Craig, E. A. (1994) *Eur. J. Biochem.* 219, 11–23.
14. Cyr, D. M., Langer, T. & Douglas, M. G. (1994) *Trends Biochem. Sci.* 19, 176–181.
15. Craig, E. A., Weissman, J. S. & Horwich, A. L. (1994) *Cell* 78, 365–372.
16. Newton, A. C. (1995) *J. Biol. Chem.* 270, 28495–28498.
17. Nishizuka, Y. (1995) *FASEB J.* 9, 484–496.
18. Rohan, P. J., Davis, P., Moskaluk, C. A., Kearns, M., Krutzsch, H., Siebenlist, U. & Kelly, K. (1993) *Science* 259, 1765–1766.
19. Thanos, D. & Maniatis, T. (1995) *Cell* 80, 529–532.
20. Baeuerle, P. A. & Henkel, T. (1994) *Annu. Rev. Immunol.* 12, 141–179.
21. Liou, H.-C. & Baltimore, D. (1993) *Curr. Opin. Cell Biol.* 5, 477–487.
22. Cohen, G. B., Ren, R. & Baltimore, D. (1995) *Cell* 80, 237–248.
23. Chan, A. C., Desai, D. M. & Weiss, A. (1994) *Annu. Rev. Immunol.* 12, 555–592.
24. Crabtree, G. R. & Clipstone, N. A. (1994) *Annu. Rev. Biochem.* 63, 1045–1083.
25. Gress, T. M., Hoheisel, J. D., Lennon, G. G., Zehetner, G. & Lehrach, H. (1992) *Mamm. Genome* 3, 609–619.
26. Bernard, K., Auphan, N., Granjeaud, S., Victorero, G., Schmitt-Verhulst, A.-M., Jordan, B. R. & Nguyen, C. (1996) *Nucleic Acids Res.* 24, 1435–1442.
27. Fodor, S. P. A., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T. & Solas, D. (1991) *Science* 251, 767–773.
28. Southern, E. M., Maskos, U. & Elder, J. K. (1992) *Genomics* 13, 1008–1017.
29. Guo, Z., Guilfoyle, R. A., Thiel, A. J., Wang, R. & Smith, L. M. (1994) *Nucleic Acids Res.* 22, 5456–5465.
30. Matson, R. S., Rampal, J., Pentoney, S. L., Jr., Anderson, P. D. & Coassin, P. (1995) *Anal. Biochem.* 224, 110–116.
31. Velculescu, V. E., Zhang, L., Vogelstein, B. & Kinzler, K. W. (1995) *Science* 270, 484–487.
32. Adams, M. D. (1996) *BioEssays* 18, 261–262.
33. Fields, C., Adams, M. D., White, O. & Venter, J. C. (1994) *Nat. Genet.* 7, 345–346.

B

S SCI08 Dup

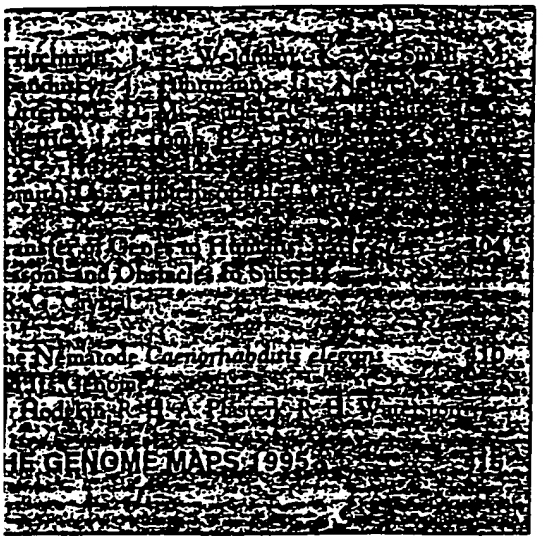
B

GENOME ISSUE

COVER

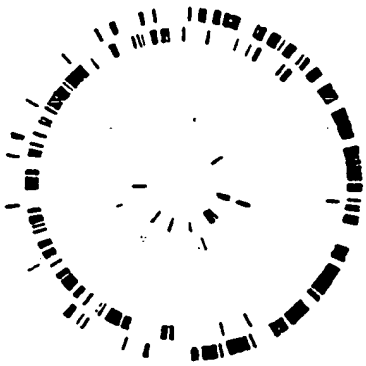
The Genome Project adds a new dimension to questions on gene expression in humans and model systems. A chart on page 415 summarizes progress in the *Caenorhabditis elegans* Genome Project and indicates some ways information about sequences can be used.

News stories, Articles, Perspectives, Policy Forums, and Reports focus on technological developments, clinical applications, and ethical concerns resulting from the burgeoning of genomic information. [*C. elegans* image: F. Maduro and D. Pilgrim, University of Alberta]



REPORTS

TECHNICAL COMMENTS



397

Good things in small genomes

◩ Indicates accompanying feature

■ SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the last week in December, by the American Association for the Advancement of Science, 1333 H Street, NW, Washington, DC 20005. Second-class postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 1995 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership and subscription (51 issues): $97 ($50 allocated to subscription). Domestic institutional subscription (51 issues): $228. Foreign postage extra: Mexico, Caribbean (surface mail) $53; other countries (air assist delivery) $93. First class, airmail, student and emeritus rates on request. Canadian rates with GST available upon request, GST #1254 88122. Printed in the U.S.A.

Change of address: allow 4 weeks, giving old and new addresses and 8-digit account number. Postmaster: Send change of address to *Science*, P.O. Box 1811, Danbury, CT 06813-1811. Single copy sales: $7.00 per issue prepaid includes surface postage; bulk rates on request. Authorization to photocopy material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that $3.00 per article is paid directly to CCC, 27 Congress Street, Salem, MA 01970. The identification code for *Science* is 0036-8075/83 $3.00. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

Axl1p sequence following Ser[208] and occurs within the domain of Axl1p that shows homology with hIDE (*14*). To delete the complete *STE23* sequence and create the *ste23Δ::URA3* mutation, polymerase chain reaction (PCR) primers (5′-TCGGAAGACCTCATTCTTGCTCATTTTGATATTGCTC- TGTAGATTGTACTGAGAGTGCAC-3′; and 5′-GCTACAAACAGCGTCGACTTGAATGCCCCGACATCTTCGACTGTGCGGTATTTCACACCG-3′) were used to amplify the *URA3* sequence of pRS316, and the reaction product was transformed into yeast for one-step gene replacement [R. Rothstein, *Methods Enzymol.* **194**, 281 (1991)]. To create the *axl1Δ::LEU2* mutation contained on p114, a 5.0-kb Sal I fragment from p*AXL1* was cloned into pUC19, and an internal 4.0-kb Hpa I–Xho I fragment was replaced with a *LEU2* fragment. To construct the *ste23Δ::LEU2* allele (a deletion corresponding to 931 amino acids) carried on p153, a *LEU2* fragment was used to replace the 2.8-kb Pml I–Ecl136 II fragment of *STE23*, which occurs within a 6.2-kb Hind III–Bgl II genomic fragment carried on pSP72 (Promega). To create YEp*MFA1*, a 1.6-kb Bam HI fragment containing *MFA1*, from pKK16 [K. Kuchler, R. E. Sterne, J. Thorner, *EMBO J.* **8**, 3973 (1989)], was ligated into the Bam HI site of YEp351 [J. E. Hill, A. M. Myers, T. J. Koerner, A. Tzagoloff, *Yeast* **2**, 163 (1986)].

24. J. Chant and I. Herskowitz, *Cell* **65**, 1203 (1991).
25. B. W. Matthews, *Acc. Chem. Res.* **21**, 333 (1988).
26. K. Kuchler, H. G. Dohlman, J. Thorner, *J. Cell Biol.* **120**, 1203 (1993); R. Kolling and C. P. Hollenberg, *EMBO J.* **13**, 3261 (1994); C. Berkower, D. Loayza, S. Michaelis, *Mol. Biol. Cell* **5**, 1185 (1994).
27. A. Bender and J. R. Pringle, *Proc. Natl. Acad. Sci. U.S.A* **86**, 9976 (1989); J. Chant, K. Corrado, J. R. Pringle, I. Herskowitz, *Cell* **65**, 1213 (1991); S. Powers, E. Gonzales, T. Christensen, J. Cubert, D. Broek, *ibid.*, p. 1225; H. O. Park, J. Chant, I. Herskowitz, *Nature* **365**, 269 (1993); J. Chant, *Trends Genet.* **10**, 328 (1994); ——— and J. R. Pringle, *J. Cell Biol.* **129**, 751 (1995); J. Chant, M. Mischke, E. Mitchell, I. Herskowitz, J. R. Pringle, *ibid.*, p. 767.
28. G. F. Sprague Jr., *Methods. Enzymol.* **194**, 77 (1991).
29. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

30. A W303 1A derivative, SY2625 (*MATa ura3-1 leu2-3, 112 trp1-1 ade2-1 can1-100 sst1Δ mfa2Δ::FUS1-lacZ his3Δ::FUS1-HIS3*), was the parent strain for the mutant search. SY2625 derivatives for the mating assays, secreted pheromone assays, and the pulse-chase experiments included the following strains: Y49 (*ste22-1*), Y115 (*mfa1Δ::LEU2*), Y142 (*axl1::URA3*), Y173 (*axl1Δ::LEU2*), Y220 (*axl1::URA3 ste23Δ::URA3*), Y221 (*ste23Δ::URA3*), Y231 (*axl1Δ::LEU2 ste23Δ::LEU2*), and Y233 (*ste23Δ::LEU2*). *MATα* derivatives of SY2625 included the following strains: Y199 (SY2625 made *MATα*), Y278 (*ste22-1*), Y195 (*mfa1Δ::LEU2*), Y196 (*axl1Δ::LEU2*), and Y197 (*axl1::URA3*). The EG123 (*MATa leu2 ura3 trp1 can1 his4*) genetic background was used to create a set of strains for analysis of bud site selection. EG123 derivatives included the following strains: Y175 (*axl1Δ::LEU2*), Y223 (*axl1::URA3*), Y234 (*ste23Δ::LEU2*), and Y272 (*axl1Δ::LEU2 ste23Δ::LEU2*). *MATα* derivatives of EG123 included the following strains: Y214 (EG123 made *MATα*) and Y293 (*axl1Δ::LEU2*). All strains were generated by means of standard genetic or molecular methods involving the appropriate constructs (*23*). In particular, the *axl1 ste23* double mutant strains were created by crossing of the appropriate *MATa ste23* and *MATα axl1* mutants, followed by sporulation of the resultant diploid and isolation of the double mutant from nonparental di-type tetrads. Gene disruptions were confirmed with either PCR or Southern (DNA) analysis.

31. p129 is a YEp352 [J. E. Hill, A. M. Myers, T. J. Koerner, A. Tzagoloff, *Yeast* **2**, 163 (1986)] plasmid containing a 5.5-kb Sal I fragment of p*AXL1*. p151 was derived from p129 by insertion of a linker at the Bgl II site within *AXL1*, which led to an in-frame insertion of the hemagglutinin (HA) epitope (DQYPYDVPDYA) (*29*) between amino acids 854 and 855 of the *AXL1* product. pC225 is a KS+ (Stratagene) plasmid containing a 0.5-kb Bam HI–Sst I fragment from p*AXL1*. Substitution mutations of the proposed active site of Axl1p were created with the use of pC225 and site-specific mutagenesis involving appropriate synthetic oligonucleotides (*axl1-H68A*, 5′-GTGCTCACAAAGCGCTGCCAAACCGGC-3′; *axl1-E71A*, 5′-AAGAATCATGTGCGCACAAAGGTGCGC-3′; and *axl1-E71D*, 5′-AAGAATCATGTGATCACAAAGGTGCGC-3′). The mutations were confirmed by sequence analysis. After mutagenesis, the 0.4-kb Bam HI–Msc I fragment from the mutagenized pC225 plasmids was transferred into p*AXL1* to create a set of pRS316 plasmids carrying different *AXL1* alleles, p124 (*axl1-H68A*), p130 (*axl1-E71A*), and p132 (*axl1-E71D*). Similarly, a set of HA-tagged alleles carried on YEp352 were created after replacement of the p151 Bam HI–Msc I fragment, to generate p161 (*axl1-E71A*), p162 (*axl1-H68A*), and p163 (*axl1-E71D*).

32. We thank J. Becker and S. Michaelis for providing a-factor antibodies; S. Michaelis for discussing unpublished results and helping with the pulse-chase experiments; J. Brown, J. Chant, and S. Sanders for their input concerning bud site selection experiments; M. Raymond, F. Tamanoi, and M. Whiteway for plasmids; M. Marra for providing the *STE23* genomic fragment; and H. Bussey, J. Brown, N. Davis, T. Favero, C. de Hoog, and S. Kim for comments on the manuscript. Supported by a grant to C.B. from the Natural Sciences and Engineering Research Council of Canada. Support for M.N.A. was from a California Tobacco-Related Disease Research Program postdoctoral fellowship (4FT-0083).

22 June 1995; accepted 21 August 1995

Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray

Mark Schena,*† Dari Shalon,*† Ronald W. Davis, Patrick O. Brown‡

A high-capacity system was developed to monitor the expression of many genes in parallel. Microarrays prepared by high-speed robotic printing of complementary DNAs on glass were used for quantitative expression measurements of the corresponding genes. Because of the small format and high density of the arrays, hybridization volumes of 2 microliters could be used that enabled detection of rare transcripts in probe mixtures derived from 2 micrograms of total cellular messenger RNA. Differential expression measurements of 45 *Arabidopsis* genes were made by means of simultaneous, two-color fluorescence hybridization.

The temporal, developmental, topographical, histological, and physiological patterns in which a gene is expressed provide clues to its biological role. The large and expanding database of complementary DNA (cDNA) sequences from many organisms (*1*) presents the opportunity of defining these patterns at the level of the whole genome.

For these studies, we used the small flowering plant *Arabidopsis thaliana* as a model organism. *Arabidopsis* possesses many advantages for gene expression analysis, including the fact that it has the smallest genome of any higher eukaryote examined to date (*2*). Forty-five cloned *Arabidopsis* cDNAs (Table 1), including 14 complete sequences and 31 expressed sequence tags (ESTs), were used as gene-specific targets. We obtained the ESTs by selecting cDNA clones at random from an *Arabidopsis* cDNA library. Sequence analysis revealed that 28 of the 31 ESTs matched sequences in the database (Table 1). Three additional cDNAs from other organisms served as controls in the experiments.

The 48 cDNAs, averaging ~1.0 kb, were amplified with the polymerase chain reaction (PCR) and deposited into individual wells of a 96-well microtiter plate. Each sample was duplicated in two adjacent wells to allow the reproducibility of the arraying and hybridization process to be tested. Samples from the microtiter plate were printed onto glass microscope slides in an area measuring 3.5 mm by 5.5 mm with the use of a high-speed arraying machine (*3*). The arrays were processed by chemical and heat treatment to attach the DNA sequences to the glass surface and denature them (*3*). Three arrays, printed in a single lot, were used for the experiments here. A single microtiter plate of PCR products provides sufficient material to print at least 500 arrays.

Fluorescent probes were prepared from total *Arabidopsis* mRNA (*4*) by a single round of reverse transcription (*5*). The *Arabidopsis* mRNA was supplemented with human acetylcholine receptor (AChR) mRNA at a dilution of 1:10,000 (w/w) before cDNA synthesis, to provide an internal standard for calibration (*5*). The resulting fluorescently labeled cDNA mixture was hybridized to an array at high stringency (*6*) and scanned

M. Schena and R. W. Davis, Department of Biochemistry, Beckman Center, Stanford University Medical Center, Stanford, CA 94305, USA.
D. Shalon and P. O. Brown, Department of Biochemistry and Howard Hughes Medical Institute, Beckman Center, Stanford University Medical Center, Stanford, CA 94305, USA.

*These authors contributed equally to this work.
†Present address: Synteni, Palo Alto, CA 94303, USA.
‡To whom correspondence should be addressed. E-mail: pbrown@cmgm.stanford.edu

with a laser (3). A high-sensitivity scan gave signals that saturated the detector at nearly all of the *Arabidopsis* target sites (Fig. 1A). Calibration relative to the AChR mRNA standard (Fig. 1A) established a sensitivity limit of −1:50,000. No detectable hybridization was observed to either the rat glucocorticoid receptor (Fig. 1A) or the yeast *TRP4* (Fig. 1A) targets even at the highest scanning sensitivity. A moderate-sensitivity scan of the same array allowed linear detection of the more abundant transcripts (Fig. 1B). Quantitation of both scans revealed a range of expression levels spanning three orders of magnitude for the 45 genes tested (Table 2). RNA blots (7) for several genes (Fig. 2) corroborated the expression levels measured with the microarray to within a factor of 5 (Table 2).

Differential gene expression was investigated with a simultaneous, two-color hybridization scheme, which served to minimize experimental variation inherent in the comparison of independent hybridizations. Fluorescent probes were prepared from two mRNA sources with the use of reverse transcriptase in the presence of fluorescein- and lissamine-labeled nucleotide analogs, respectively (5). The two probes were then mixed together in equal proportions, hybridized to a single array, and scanned separately for fluorescein and lissamine emission after independent excitation of the two fluorophores (3).

To test whether overexpression of a single gene could be detected in a pool of total *Arabidopsis* mRNA, we used a microarray to analyze a transgenic line overexpressing the single transcription factor *HAT4* (8). Fluorescent probes representing mRNA from wild-type and *HAT4*-transgenic plants were labeled with fluorescein and lissamine, respectively; the two probes were then mixed and hybridized to a single array. An intense hybridization signal was observed at the position of the *HAT4* cDNA in the lissamine-specific scan (Fig. 1D), but not in the fluorescein-specific scan of the same array (Fig. 1C). Calibration with AChR mRNA added to the fluorescein and lissamine cDNA synthesis reactions at dilutions of 1:10,000 (Fig. 1C) and 1:100 (Fig. 1D), respectively, revealed a 50-fold elevation of *HAT4* mRNA in the transgenic line relative to its abundance in wild-type plants (Table 2). This magnitude of *HAT4* overexpression matched that inferred from the Northern (RNA) analysis within a factor of 2 (Fig. 2 and Table 2). Expression of all the other genes monitored on the array differed by less than a factor of 5 between *HAT4*-transgenic and wild-type plants (Fig 1, C

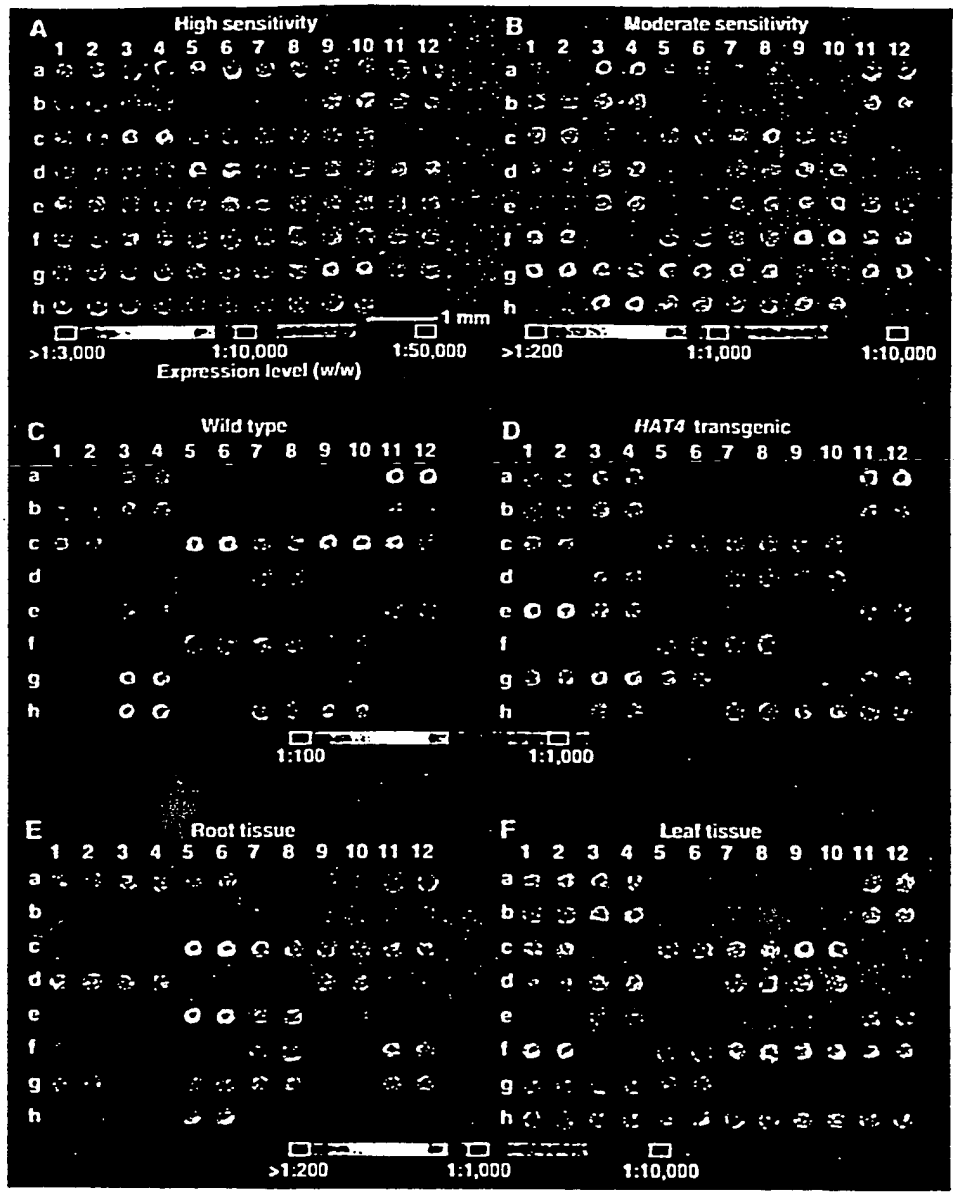


Fig. 1. Gene expression monitored with the use of cDNA microarrays. Fluorescent scans represented in pseudocolor correspond to hybridization intensities. Color bars were calibrated from the signal obtained with the use of known concentrations of human AChR mRNA in independent experiments. Numbers and letters on the axes mark the position of each cDNA. (A) High-sensitivity fluorescein scan after hybridization with fluorescein-labeled cDNA derived from wild-type plants. (B) Same array as in (A) but scanned at moderate sensitivity. (C and D) A single array was probed with a 1:1 mixture of fluorescein-labeled cDNA from wild-type plants and lissamine-labeled cDNA from HAT4-transgenic plants. The single array was then scanned successively to detect the fluorescein fluorescence corresponding to mRNA from wild-type plants (C) and the lissamine fluorescence corresponding to mRNA from HAT4-transgenic plants (D). (E and F) A single array was probed with a 1:1 mixture of fluorescein-labeled cDNA from root tissue and lissamine-labeled cDNA from leaf tissue. The single array was then scanned successively to detect the fluorescein fluorescence corresponding to mRNAs expressed in roots (E) and the lissamine fluorescence corresponding to mRNAs expressed in leaves (F).

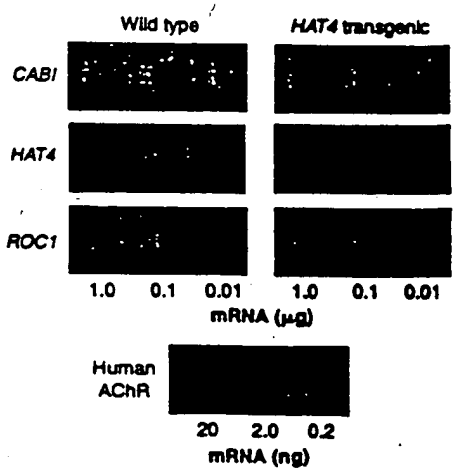


Fig. 2. Gene expression monitored with RNA (Northern) blot analysis. Designated amounts of mRNA from wild-type and *HAT4*-transgenic plants were spotted onto nylon membranes and probed with the cDNAs indicated. Purified human AChR mRNA was used for calibration.

and D, and Table 2). Hybridization of fluorescein-labeled glucocorticoid receptor cDNA (Fig. 1C) and lissamine-labeled *TRP4* cDNA (Fig. 1D) verified the presence of the negative control targets and the lack of optical cross talk between the two fluorophores.

To explore a more complex alteration in expression patterns, we performed a second two-color hybridization experiment with fluorescein- and lissamine-labeled probes prepared from root and leaf mRNA, respectively. The scanning sensitivities for the two fluorophores were normalized by matching the signals resulting from AChR mRNA, which was added to both cDNA synthesis reactions at a dilution of 1:1000 (Fig. 1, E and F). A comparison of the scans revealed widespread differences in gene expression between root and leaf tissue (Fig. 1, E and F). The mRNA from the light-regulated *CAB1* gene was ~500-fold more abundant in leaf (Fig. 1F) than in root tissue (Fig. 1E). The expression of 26 other genes differed between root and leaf tissue by more than a factor of 5 (Fig. 1, E and F).

The *HAT4*-transgenic line we examined has elongated hypocotyls, early flowering, poor germination, and altered pigmentation (*8*). Although changes in expression were observed for *HAT4*, large changes in expression were not observed for any of the other 44 genes we examined. This was somewhat surprising, particularly because comparative analysis of leaf and root tissue identified 27 differentially expressed genes. Analysis of an expanded set of genes may be required to identify genes whose expression changes upon *HAT4* overexpression; alternatively, a comparison of mRNA populations from specific tissues of wild-type and *HAT4*-transgenic plants may allow identification of downstream genes.

At the current density of robotic printing, it is feasible to scale up the fabrication process to produce arrays containing 20,000 cDNA targets. At this density, a single array would be sufficient to provide gene-specific targets encompassing nearly the entire repertoire of expressed genes in the *Arabidopsis* genome (*2*). The availability of 20,274 ESTs from *Arabidopsis* (*1*, *9*) would provide a rich source of templates for such studies.

The estimated 100,000 genes in the human genome (*10*) exceeds the number of *Arabidopsis* genes by a factor of 5 (*2*). This modest increase in complexity suggests that similar cDNA microarrays, prepared from the rapidly growing repertoire of human ESTs (*1*), could be used to determine the expression patterns of tens of thousands of human genes in diverse cell types. Coupling an amplification strategy to the reverse transcription reaction (*11*) could make it feasible to monitor expression even in minute tissue samples. A wide variety of acute and chronic physiological and pathological conditions might lead to characteristic changes in the patterns of gene expression in peripheral blood cells or other easily sampled tissues. In concert with cDNA microarrays for monitoring complex expression patterns, these tissues might therefore serve as sensitive in vivo sensors for clinical diagnosis. Microarrays of cDNAs could thus provide a useful link between human gene sequences and clinical medicine.

Table 1. Sequences contained on the cDNA microarray. Shown is the position, the known or putative function, and the accession number of each cDNA in the microarray (Fig. 1). All but three of the ESTs used in this study matched a sequence in the database. NADH, reduced form of nicotinamide adenine dinucleotide; ATPase, adenosine triphosphatase; GTP, guanosine triphosphate.

Position	cDNA	Function	Accession number
a1, 2	AChR	Human AChR	*
a3, 4	EST3	Actin	H36236
a5, 6	EST6	NADH dehydrogenase	Z27010
a7, 8	*AAC1*	Actin 1	M20016
a9, 10	EST12	Unknown	U36594†
a11, 12	EST13	Actin	T45783
b1, 2	*CAB1*	Chlorophyll a/b binding	M85150
b3, 4	EST17	Phosphoglycerate kinase	T44490
b5, 6	*GA4*	Gibberellic acid biosynthesis	L37126
b7, 8	EST19	Unknown	U36595†
b9, 10	*GBF-1*	G-box binding factor 1	X63894
b11, 12	EST23	Elongation factor	X52256
c1, 2	EST29	Aldolase	T04477
c3, 4	*GBF-2*	G-box binding factor 2	X63895
c5, 6	EST34	Chloroplast protease	R87034
c7, 8	EST35	Unknown	T14152
c9, 10	EST41	Catalase	T22720
c11, 12	rGR	Rat glucocorticoid receptor	M14053
d1, 2	EST42	Unknown	U36596†
d3, 4	EST45	ATPase	J04185
d5, 6	*HAT1*	Homeobox-leucine zipper 1	U09332
d7, 8	EST46	Light harvesting complex	T04063
d9, 10	EST49	Unknown	T76267
d11, 12	*HAT2*	Homeobox-leucine zipper 2	U09335
e1, 2	*HAT4*	Homeobox-leucine zipper 4	M90394
e3, 4	EST50	Phosphoribulokinase	T04344
e5, 6	*HAT5*	Homeobox-leucine zipper 5	M90416
e7, 8	EST51	Unknown	Z33675
e9, 10	*HAT22*	Homeobox-leucine zipper 22	U09336
e11, 12	EST52	Oxygen evolving	T21749
f1, 2	EST59	Unknown	Z34607
f3, 4	*KNAT1*	*Knotted*-like homeobox 1	U14174
f5, 6	EST60	RuBisCO small subunit	X14564
f7, 8	EST69	Translation elongation factor	T42799
f9, 10	*PPH1*	Protein phosphatase 1	U34803
f11, 12	EST70	Unknown	T44621
g1, 2	EST75	Chloroplast protease	T43698
g3, 4	EST78	Unknown	R65481
g5, 6	*ROC1*	Cyclophilin	L14844
g7, 8	EST82	GTP binding	X59152
g9, 10	EST83	Unknown	Z33795
g11, 12	EST84	Unknown	T45278
h1, 2	EST91	Unknown	T13832
h3, 4	EST96	Unknown	R64816
h5, 6	*SAR1*	Synaptobrevin	M90418
h7, 8	EST100	Light harvesting complex	Z18205
h9, 10	EST103	Light harvesting complex	X03909
h11, 12	*TRP4*	Yeast tryptophan biosynthesis	X04273

*Proprietary sequence of Stratagene (La Jolla, California). †No match in the database; novel EST.

Table 2. Gene expression monitoring by microarray and RNA blot analyses; tg, *HAT4*-transgenic. See Table 1 for additional gene information. Expression levels (w/w) were calibrated with the use of known amounts of human AChR mRNA. Values for the microarray were determined from microarray scans (Fig. 1); values for the RNA blot were determined from RNA blots (Fig. 2).

Gene	Expression level (w/w)	
	Microarray	RNA blot
CAB1	1:48	1:83
CAB1 (tg)	1:120	1:150
HAT4	1:8300	1:6300
HAT4 (tg)	1:150	1:210
ROC1	1:1200	1:1800
ROC1 (tg)	1:260	1:1300

REFERENCES AND NOTES

urrent EST database (dbEST release 091495) the National Center for Biotechnology Informa- Bethesda, MD) contains a total of 322,225 en- including 255,645 from the human genome :1,044 from *Arabidopsis*. Access is available via /orld Wide Web (http://www.ncbi.nlm.nih.gov).
Meyerowitz and R. E. Pruitt, *Science* 229, 1214 i); R. E. Pruitt and E. M. Meyerowitz, *J. Mol. Biol.* 169 (1986); I. Hwang *et al.*, *Plant J.* 1, 367 (1991); rvis *et al.*, *Plant Mol. Biol.* 24, 685 (1994); L. Le *i et al.*, *Mol. Gen. Genet.* 245, 390 (1994).
ialon, thesis, Stanford University (1995); ——— P. O. Brown, in preparation. Microarrays were ated on poly-L-lysine–coated microscope s (Sigma) with a custom-built arraying machine with one printing tip. The tip loaded 1 μl of PCR uct (0.5 mg/ml) from 96-well microtiter plates deposited ~0.005 μl per slide on 40 slides at a ng of 500 μm. The printed slides were rehydrat- r 2 hours in a humid chamber, snap-dried at C for 1 min, rinsed in 0.1% SDS, and treated 0.05% succinic anhydride prepared in buffer sting of 50% 1-methyl-2-pyrrolidinone and bonic acid. The cDNA on the slides was dena- in distilled water for 2 min at 90°C immediately e use. Microarrays were scanned with a laser scent scanner that contained a computer-con- l XY stage and a microscope objective. A mixed multiline laser allowed sequential excitation of vo fluorophores. Emitted light was split accord- wavelength and detected with two photomul- tubes. Signals were read into a PC with the use 2-bit analog-to-digital board. Additional details roarray fabrication and use may be obtained by s of e-mail (pbrown@cmgm.stanford.edu).
Ausubel *et al.*, Eds., *Current Protocols in Mo- r Biology* (Greene & Wiley Interscience, New 1994), pp. 4.3.1–4.3.4.
denylated [poly(A)$^+$] mRNA was prepared from RNA with the use of Oligotex-dT resin (Qiagen). se transcription (RT) reactions were carried out StrataScript RT-PCR kit (Stratagene) modified lows: 50-μl reactions contained 0.1 μg/μl of *dopsis* mRNA, 0.1 ng/μl of human AChR , 0.05 μg/μl of oligo(dT) (21-mer), 1× first l buffer, 0.03 U/μl of ribonuclease block, 500 eoxyadenosine triphosphate (dATP), 500 μM guanosine triphosphate, 500 μM dTTP, 40 eoxycytosine triphosphate (dCTP), 40 μM flu- ein-12-dCTP (or lissamine-5-dCTP), and 0.03 l StrataScript reverse transcriptase. Reactions ncubated for 60 min at 37°C, precipitated with ol, and resuspended in 10 μl of TE (10 mM tris- nd 1 mM EDTA, pH 8.0). Samples were then d for 3 min at 94°C and chilled on ice. The RNA legraded by adding 0.25 μl of 10 N NaOH ed by a 10-min incubation at 37°C. The sam- were neutralized by addition of 2.5 μl of 1 M (pH 8.0) and 0.25 μl of 10 N HCl and precip- with ethanol. Pellets were washed with 70% ol, dried to completion in a speedvac, resus- d in 10 μl of H_2O, and reduced to 3.0 μl in a ivac. Fluorescent nucleotide analogs were ob- l from New England Nuclear (DuPont).
dization reactions contained 1.0 μl of fluorescent synthesis product (5) and 1.0 μl of hybridization [10× saline sodium citrate (SSC) and 0.2% The 2.0-μl probe mixtures were aliquoted onto icroarray surface and covered with cover slips m round). Arrays were transferred to a hybrid- chamber (3) and incubated for 18 hours at Arrays were washed for 5 min at room temper- 25°C) in low-stringency wash buffer (1× SSC 1% SDS), then for 10 min at room temperature h-stringency wash buffer (0.1× SSC and 0.1% Arrays were scanned in 0.1× SSC with the use ioresecence laser-scanning device (3).
es of poly(A)$^+$ mRNA (4, 5) were spotted onto membranes (Nytran) and crosslinked with ul- et light with the use of a Stratalinker 1800 gene). Probes were prepared by random g with the use of a Prime-It II kit (Stratagene) in sence of [^{32}P]dATP. Hybridizations were car- ut according to the instructions of the manu- facturer. Quantitation was performed on a Phosphorimager (Molecular Dynamics).
8. M. Schena and R. W. Davis, *Proc. Natl. Acad. Sci. U.S.A.* 89, 3894 (1992); M. Schena, A. M. Lloyd, R. W. Davis, *Genes Dev.* 7, 367 (1993); M. Schena and R. W. Davis, *Proc. Natl. Acad. Sci. U.S.A.* 91, 8393 (1994).
9. H. Hofte *et al.*, *Plant J.* 4, 1051 (1993); T. Newman *et al.*, *Plant Physiol.* 106, 1241 (1994).
10. N. E. Morton, *Proc. Natl. Acad. Sci. U.S.A.* 88, 7474 (1991); E. D. Green and R. H. Waterston, *J. Am. Med. Assoc.* 266, 1966 (1991); C. Betarme-Charte- lot, *Cell* 70, 1059 (1992); D. R. Cox *et al.*, *Science* 265, 2031 (1994).
11. E. S. Kawasaki *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 85, 5698 (1988).
12. The laser fluorescent scanner was designed and fabricated in collaboration with S. Smith of Stanford University. Scanner and analysis software was developed by R. X. Xia. The succinic anhydride reaction was suggested by J. Mulligan and J. Van Ness of Darwin Molecular Corporation. Thanks to S. Theologis, C. Somerville, K. Yamamoto, and members of the laboratories of R.W.D. and P.O.B. for critical comments. Supported by the Howard Hughes Medical Institute and by grants from NIH (R21HG00450) (P.O.B.) and R37AG00198 (R.W.D.)) and from NSF (MCB9106011) (R.W.D.) and by an NSF graduate fellowship (D.S.). P.O.B. is an assistant investigator of the Howard Hughes Medical Institute.

11 August 1995; accepted 22 September 1995

Gene Therapy in Peripheral Blood Lymphocytes and Bone Marrow for ADA$^-$ Immunodeficient Patients

Claudio Bordignon,* Luigi D. Notarangelo, Nadia Nobili, Giuliana Ferrari, Giulia Casorati, Paola Panina, Evelina Mazzolari, Daniela Maggioni, Claudia Rossi, Paolo Servida, Alberto G. Ugazio, Fulvio Mavilio

Adenosine deaminase (ADA) deficiency results in severe combined immunodeficiency, the first genetic disorder treated by gene therapy. Two different retroviral vectors were used to transfer ex vivo the human ADA minigene into bone marrow cells and peripheral blood lymphocytes from two patients undergoing exogenous enzyme replacement therapy. After 2 years of treatment, long-term survival of T and B lymphocytes, marrow cells, and granulocytes expressing the transferred ADA gene was demonstrated and resulted in normalization of the immune repertoire and restoration of cellular and humoral immunity. After discontinuation of treatment, T lymphocytes, derived from transduced peripheral blood lymphocytes, were progressively replaced by marrow-derived T cells in both patients. These results indicate successful gene transfer into long-lasting progenitor cells, producing a functional multilineage progeny.

Severe combined immunodeficiency associated with inherited deficiency of ADA (*1*) is usually fatal unless affected children are kept in protective isolation or the immune system is reconstituted by bone marrow transplantation from a human leukocyte antigen (HLA)–identical sibling donor (*2*). This is the therapy of choice, although it is available only for a minority of patients. In recent years, other forms of therapy have been developed, including transplants from haploidentical donors (*3*, *4*), exogenous enzyme replacement (*5*), and somatic-cell gene therapy (*6*–*9*).

We previously reported a preclinical model in which ADA gene transfer and expression successfully restored immune functions in human ADA-deficient (ADA$^-$) peripheral blood lymphocytes (PBLs) in immunodeficient mice in vivo (*10*, *11*). On the basis of these preclinical results, the clinical application of gene therapy for the treatment of ADA$^-$ SCID (severe combined immunodeficiency disease) patients who previously failed exogenous enzyme replacement therapy was approved by our Institutional Ethical Committees and by the Italian National Committee for Bioethics (*12*). In addition to evaluating the safety and efficacy of the gene therapy procedure, the aim of the study was to define the relative role of PBLs and hematopoietic stem cells in the long-term reconstitution of immune functions after retroviral vector–mediated ADA gene transfer. For this purpose, two structurally identical vectors expressing the human ADA complementary DNA (cDNA), distinguishable by the presence of alternative restriction sites in a nonfunctional region of the viral long-terminal repeat (LTR), were used to transduce PBLs and bone marrow (BM) cells independently. This procedure allowed identification of the origin of

C. Bordignon, N. Nobili, G. Ferrari, D. Maggioni, C. Rossi, P. Servida, F. Mavilio, Telethon Gene Therapy Program for Genetic Diseases, DIBIT, Istituto Scientifico H. S. Raffaele, Milan, Italy.
L. D. Notarangelo, E. Mazzolari, A. G. Ugazio, Department of Pediatrics, University of Brescia Medical School, Brescia, Italy.
G. Casorati, Unità di Immunochimica, DIBIT, Istituto Scientifico H. S. Raffaele, Milan, Italy.
P. Panina, Roche Milano Ricerche, Milan, Italy.
*To whom correspondence should be addressed.

C

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G01N 33/543, 33/68 — A1

(11) International Publication Number: WO 95/35505

(43) International Publication Date: 28 December 1995 (28.12.95)

(21) International Application Number: PCT/US95/07659

(22) International Filing Date: 16 June 1995 (16.06.95)

(30) Priority Data:

08/261,388	17 June 1994 (17.06.94)	US
08/477,809	7 June 1995 (07.06.95)	US

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(71) Applicant: THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY [US/US]; Stanford, CA 94305 (US).

(72) Inventors: SHALON, Tidhar, Dari; 364 Fletcher Drive, Atherton, CA 94027 (US). BROWN, Patrick, O.; 76 Peter Coutts Circle, Stanford, CA 94305 (US).

(74) Agent: DEHLINGER, Peter, J.; Dehlinger & Associates, P.O. Box 60850, Palo Alto, CA 94306-1546 (US).

(54) Title: METHOD AND APPARATUS FOR FABRICATING MICROARRAYS OF BIOLOGICAL SAMPLES

(57) Abstract

A method and apparatus for forming microarrays of biological samples on a support are disclosed. The method involves dispensing a known volume of a reagent at each of a selected array position, by tapping a capillary dispenser on the support under conditions effective to draw a defined volume of liquid onto the support. The apparatus is designed to produce a microarray of such regions in an automated fashion.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

METHOD AND APPARATUS FOR FABRICATING MICROARRAYS OF BIOLOGICAL SAMPLES

Field of the Invention

This invention relates to a method and apparatus for fabricating microarrays of biological samples for large scale screening assays, such as arrays of DNA samples to be used in DNA hybridization assays for genetic research and diagnostic applications.

References

Abouzied, *et al.*, *Journal of AOAC International* 77(2):495-500 (1994).

Bohlander, *et al.*, *Genomics* 13:1322-1324 (1992).

Drmanac, *et al.*, *Science* 260:1649-1652 (1993).

Fodor, *et al.*, *Science* 251:767-773 (1991).

Khrapko, *et al.*, *DNA Sequence* 1:375-388 (1991).

Kuriyama, *et al.*, AN ISFET BIOSENSOR, APPLIED BIOSENSORS (Donald Wise, Ed.), Butterworths, pp. 93-114 (1989).

Lehrach, *et al.*, HYBRIDIZATION FINGERPRINTING IN GENOME MAPPING AND SEQUENCING, GENOME ANALYSIS, VOL 1 (Davies and Tilgham, Eds.), Cold Spring Harbor Press, pp. 39-81 (1990).

Maniatis, *et al.*, MOLECULAR CLONING, A LABORATORY MANUAL, Cold Spring Harbor Press (1989).

Nelson, *et al.*, *Nature Genetics* 4:11-18 (1993).

Pirrung, *et al.*, U.S. Patent No. 5,143,854 (1992).

Riles, *et al.*, *Genetics* 134:81-150 (1993).

Schena, M. *et al.*, *Proc. Nat. Acad. Sci. USA* 89:3894-3898 (1992).

Southern, *et al.*, *Genomics* 13:1008-1017 (1992).

Background of the Invention

A variety of methods are currently available for making arrays of biological macromolecules, such as arrays of nucleic acid molecules or proteins. One method for making ordered arrays of DNA on a porous membrane is a "dot blot" approach. In this method, a vacuum manifold transfers a plurality, *e.g.*, 96, aqueous samples of DNA from 3 millimeter diameter wells to a porous membrane. A common variant of this procedure is a "slot-blot" method in which the wells have highly-elongated oval shapes.

The DNA is immobilized on the porous membrane by baking the membrane or exposing it to UV radiation. This is a manual procedure practical for making one array at a time and usually limited to 96 samples per array. "Dot-blot" procedures are therefore inadequate for applications in which many thousand samples must be determined.

A more efficient technique employed for making ordered arrays of genomic fragments uses an array of pins dipped into the wells, *e.g.*, the 96 wells of a microtitre plate, for transferring an array of samples to a substrate, such as a porous membrane. One array includes pins that are designed to spot a membrane in a staggered fashion, for creating an array of 9216 spots in a 22 × 22 cm area (Lehrach, *et al.*, 1990). A limitation with this approach is that the volume of DNA spotted in each pixel of each array is highly variable.

In addition, the number of arrays that can be made with each dipping is usually quite small.

An alternate method of creating ordered arrays of nucleic acid sequences is described by Pirrung, *et al.* (1992), and also by Fodor, *et al.* (1991). The method involves synthesizing different nucleic acid sequences at different discrete regions of a support. This method employs elaborate synthetic schemes, and is generally limited to relatively short nucleic acid sample, *e.g.*, less than 20 bases. A related method has been described by Southern, *et al.* (1992).

Khrapko, *et al.* (1991) describes a method of making an oligonucleotide matrix by spotting DNA onto a thin layer of polyacrylamide. The spotting is done manually with a micropipette.

None of the methods or devices described in the prior art are designed for mass fabrication of microarrays characterized by (i) a large number of micro-sized assay regions separated by a distance of 50-200 microns or less, and (ii) a well-defined amount, typically in the picomole range, of analyte associated with each region of the array.

Furthermore, current technology is directed at performing such assays one at a time to a single array of DNA molecules. For example, the most common method for performing DNA hybridizations to arrays spotted onto porous membrane involves sealing the membrane in a plastic bag (Maniatas, *et al.*, 1989) or a rotating glass cylinder (Robbins Scientific) with the labeled hybridization probe inside the sealed chamber. For arrays made on non-porous surfaces, such as a microscope slide, each array is incubated with the labeled hybridization probe sealed under a coverslip. These techniques require a separate sealed chamber for

each array which makes the screening and handling of many such arrays inconvenient and time intensive.

Abouzied, *et al.* (1994) describes a method of printing horizontal lines of antibodies on a nitrocellulose membrane and separating regions of the membrane with vertical stripes of a hydrophobic material. Each vertical stripe is then reacted with a different antigen and the reaction between the immobilized antibody and an antigen is detected using a standard ELISA colorimetric technique. Abouzied's technique makes it possible to screen many one-dimensional arrays simultaneously on a single sheet of nitrocellulose. Abouzied makes the nitrocellulose somewhat hydrophobic using a line drawn with PAP Pen (Research Products International). However Abouzied does not describe a technology that is capable of completely sealing the pores of the nitrocellulose. The pores of the nitrocellulose are still physically open and so the assay reagents can leak through the hydrophobic barrier during extended high temperature incubations or in the presence of detergents which makes the Abouzied technique unacceptable for DNA hybridization assays.

Porous membranes with printed patterns of hydrophilic/hydrophobic regions exist for applications such as ordered arrays of bacteria colonies. QA Life Sciences (San Diego CA) makes such a membrane with a grid pattern printed on it. However, this membrane has the same disadvantage as the Abouzied technique since reagents can still flow between the gridded arrays making them unusable for separate DNA hybridization assays.

Pall Corporation make a 96-well plate with a porous filter heat sealed to the bottom of the plate. These plates are capable of containing different

reagents in each well without cross-contamination. However, each well is intended to hold only one target element whereas the invention described here makes a microarray of many biomolecules in each subdivided region of the solid support. Furthermore, the 96 well plates are at least 1 cm thick and prevent the use of the device for many colorimetric, fluorescent and radioactive detection formats which require that the membrane lie flat against the detection surface. The invention described here requires no further processing after the assay step since the barriers elements are shallow and do not interfere with the detection step thereby greatly increasing convenience.

Hyseq Corporation has described a method of making an "array of arrays" on a non-porous solid support for use with their sequencing by hybridization technique. The method described by Hyseq involves modifying the chemistry of the solid support material to form a hydrophobic grid pattern where each subdivided region contains a microarray of biomolecules. Hyseq's flat hydrophobic pattern does not make use of physical blocking as an additional means of preventing cross contamination.

Summary of the Invention

The invention includes, in one aspect, a method of forming a microarray of analyte-assay regions on a solid support, where each region in the array has a known amount of a selected, analyte-specific reagent. The method involves first loading a solution of a selected analyte-specific reagent in a reagent-dispensing device having an elongate capillary channel (i) formed by spaced-apart, coextensive elongate members, (ii) adapted to hold a quantity of the reagent solution and (iii) having a tip region at which aqueous

solution in the channel forms a meniscus. The channel is preferably formed by a pair of spaced-apart tapered elements.

The tip of the dispensing device is tapped against a solid support at a defined position on the support surface with an impulse effective to break the meniscus in the capillary channel deposit a selected volume of solution on the surface, preferably a selected volume in the range 0.01 to 100 nl. The two steps are repeated until the desired array is formed.

The method may be practiced in forming a plurality of such arrays, where the solution-depositing step is are applied to a selected position on each of a plurality of solid supports at each repeat cycle.

The dispensing device may be loaded with a new solution, by the steps of (i) dipping the capillary channel of the device in a wash solution, (ii) removing wash solution drawn into the capillary channel, and (iii) dipping the capillary channel into the new reagent solution.

Also included in the invention is an automated apparatus for forming a microarray of analyte-assay regions on a plurality of solid supports, where each region in the array has a known amount of a selected, analyte-specific reagent. The apparatus has a holder for holding, at known positions, a plurality of planar supports, and a reagent dispensing device of the type described above.

The apparatus further includes positioning structure for positioning the dispensing device at a selected array position with respect to a support in said holder, and dispensing structure for moving the dispensing device into tapping engagement against a support with a selected impulse effective to deposit a

selected volume on the support, e.g., a selected volume in the volume range 0.01 to 100 nl.

The positioning and dispensing structures are controlled by a control unit in the apparatus. The unit operates to (i) place the dispensing device at a loading station, (ii) move the capillary channel in the device into a selected reagent at the loading station, to load the dispensing device with the reagent, and (iii) dispense the reagent at a defined array position on each of the supports on said holder. The unit may further operate, at the end of a dispensing cycle, to wash the dispensing device by (i) placing the dispensing device at a washing station, (ii) moving the capillary channel in the device into a wash fluid, to load the dispensing device with the fluid, and (iii) remove the wash fluid prior to loading the dispensing device with a fresh selected reagent.

The dispensing device in the apparatus may be one of a plurality of such devices which are carried on the arm for dispensing different analyte assay reagents at selected spaced array positions.

In another aspect, the invention includes a substrate with a surface having a microarray of at least 10^3 distinct polynucleotide or polypeptide biopolymers in a surface area of less than about 1 cm^2. Each distinct biopolymer (i) is disposed at a separate, defined position in said array, (ii) has a length of at least 50 subunits, and (iii) is present in a defined amount between about 0.1 femtomoles and 100 nanomoles.

In one embodiment, the surface is glass slide surface coated with a polycationic polymer, such as polylysine, and the biopolymers are polynucleotides. In another embodiment, the substrate has a water-impermeable backing, a water-permeable film formed on

the backing, and a grid formed on the film. The grid is composed of intersecting water-impervious grid elements extending from said backing to positions raised above the surface of said film, and partitions the film into a plurality of water-impervious cells. A biopolymer array is formed within each well.

More generally, there is provided a substrate for use in detecting binding of labeled polynucleotides to one or more of a plurality different-sequence, immobilized polynucleotides. The substrate includes, in one aspect, a glass support, a coating of a polycationic polymer, such as polylysine, on said surface of the support, and an array of distinct polynucleotides electrostatically bound non-covalently to said coating, where each distinct biopolymer is disposed at a separate, defined position in a surface array of polynucleotides.

In another aspect, the substrate includes a water-impermeable backing, a water-permeable film formed on the backing, and a grid formed on the film, where the grid is composed of intersecting water-impervious grid elements extending from the backing to positions raised above the surface of the film, forming a plurality of cells. A biopolymer array is formed within each cell.

Also forming part of the invention is a method of detecting differential expression of each of a plurality of genes in a first cell type, with respect to expression of the same genes in a second cell type. In practicing the method, there is first produced fluorescent-labeled cDNA's from mRNA's isolated from the two cells types, where the cDNA'S from the first and second cells are labeled with first and second different fluorescent reporters.

A mixture of the labeled cDNA's from the two cell types is added to an array of polynucleotides

representing a plurality of known genes derived from the two cell types, under conditions that result in hybridization of the cDNA's to complementary-sequence polynucleotides in the array. The array is then examined by fluorescence under fluorescence excitation conditions in which (i) polynucleotides in the array that are hybridized predominantly to cDNA's derived from one of the first and second cell types give a distinct first or second fluorescence emission color, respectively, and (ii) polynucleotides in the array that are hybridized to substantially equal numbers of cDNA's derived from the first and second cell types give a distinct combined fluorescence emission color, respectively. The relative expression of known genes in the two cell types can then be determined by the observed fluorescence emission color of each spot.

These and other objects and features of the invention will become more fully apparent when the following detailed description of the invention is read in conjunction with the accompanying figures.

Brief Description of the Drawings

Fig. 1 is a side view of a reagent-dispensing device having a open-capillary dispensing head constructed for use in one embodiment of the invention;

Figs. 2A-2C illustrate steps in the delivery of a fixed-volume bead on a hydrophobic surface employing the dispensing head from Fig. 1, in accordance with one embodiment of the method of the invention;

Fig. 3 shows a portion of a two-dimensional array of analyte-assay regions constructed according to the method of the invention;

Fig. 4 is a planar view showing components of an automated apparatus for forming arrays in accordance with the invention.

Fig. 5 shows a fluorescent image of an actual 20 x 20 array of 400 fluorescently-labeled DNA samples immobilized on a poly-l-lysine coated slide, where the total area covered by the 400 element array is 16 square millimeters;

Fig. 6 is a fluorescent image of a 1.8 cm x 1.8 cm microarray containing lambda clones with yeast inserts, the fluorescent signal arising from the hybridization to the array with approximately half the yeast genome labeled with a green fluorophore and the other half with a red fluorophore;

Fig. 7 shows the translation of the hybridization image of Fig. 6 into a karyotype of the yeast genome, where the elements of Fig.-6 microarray contain yeast DNA sequences that have been previously physically mapped in the yeast genome;

Fig. 8 show a fluorescent image of a 0.5 cm x 0.5 cm microarray of 24 cDNA clones, where the microarray was hybridized simultaneously with total cDNA from wild type *Arabidopsis* plant labeled with a green fluorophore and total cDNA from a transgenic *Arabidopsis* plant labeled with a red fluorophore, and the arrow points to the cDNA clone representing the gene introduced into the transgenic *Arabidopsis* plant;

Fig. 9 shows a plan view of substrate having an array of cells formed by barrier elements in the form of a grid;

Fig. 10 shows an enlarged plan view of one of the cells in the substrate in Fig. 9, showing an array of polynucleotide regions in the cell;

Fig. 11 is an enlarged sectional view of the substrate in Fig. 9, taken along a section line in that figure; and

Fig. 12 is a scanned image of a 3 cm x 3 cm nitrocellulose solid support containing four identical

arrays of M13 clones in each of four quadrants, where each quadrant was hybridized simultaneously to a different oligonucleotide using an open face hybridization method.

Detailed Description of the Invention

I. Definitions

Unless indicated otherwise, the terms defined below have the following meanings:

"Ligand" refers to one member of a ligand/anti-ligand binding pair. The ligand may be, for example, one of the nucleic acid strands in a complementary, hybridized nucleic acid duplex binding pair; an effector molecule in an effector/receptor binding pair; or an antigen in an antigen/antibody or antigen/antibody fragment binding pair.

"Antiligand" refers to the opposite member of a ligand/anti-ligand binding pair. The antiligand may be the other of the nucleic acid strands in a complementary, hybridized nucleic acid duplex binding pair; the receptor molecule in an effector/receptor binding pair; or an antibody or antibody fragment molecule in antigen/antibody or antigen/antibody fragment binding pair, respectively.

"Analyte" or "analyte molecule" refers to a molecule, typically a macromolecule, such as a polynucleotide or polypeptide, whose presence, amount, and/or identity are to be determined. The analyte is one member of a ligand/anti-ligand pair.

"Analyte-specific assay reagent" refers to a molecule effective to bind specifically to an analyte molecule. The reagent is the opposite member of a ligand/anti-ligand binding pair.

An "array of regions on a solid support" is a linear or two-dimensional array of preferably discrete

regions, each having a finite area, formed on the surface of a solid support.

A "microarray" is an array of regions having a density of discrete regions of at least about 100/cm^2, and preferably at least about 1000/cm^2. The regions in a microarray have typical dimensions, e.g., diameters, in the range of between about 10-250 μm, and are separated from other regions in the array by about the same distance.

A support surface is "hydrophobic" if a aqueous-medium droplet applied to the surface does not spread out substantially beyond the area size of the applied droplet. That is, the surface acts to prevent spreading of the droplet applied to the surface by hydrophobic interaction with the droplet.

A "meniscus" means a concave or convex surface that forms on the bottom of a liquid in a channel as a result of the surface tension of the liquid.

"Distinct biopolymers", as applied to the biopolymers forming a microarray, means an array member which is distinct from other array members on the basis of a different biopolymer sequence, and/or different concentrations of the same or distinct biopolymers, and/or different mixtures of distinct or different-concentration biopolymers. Thus an array of "distinct polynucleotides" means an array containing, as its members, (i) distinct polynucleotides, which may have a defined amount in each member, (ii) different, graded concentrations of given-sequence polynucleotides, and/or (iii) different-composition mixtures of two or more distinct polynucleotides.

"Cell type" means a cell from a given source, e.g., a tissue, or organ, or a cell in a given state of

differentiation, or a cell associated with a given pathology or genetic makeup.

II. Method of Microarray Formation

This section describes a method of forming a microarray of analyte-assay regions on a solid support or substrate, where each region in the array has a known amount of a selected, analyte-specific reagent.

Fig. 1 illustrates, in a partially schematic view, a reagent-dispensing device 10 useful in practicing the method. The device generally includes a reagent dispenser 12 having an elongate open capillary channel 14 adapted to hold a quantity of the reagent solution, such as indicated at 16, as will be described below. The capillary channel is formed by a pair of spaced-apart, coextensive, elongate members 12a, 12b which are tapered toward one another and converge at a tip or tip region 18 at the lower end of the channel. More generally, the open channel is formed by at least two elongate, spaced-apart members adapted to hold a quantity of reagent solutions and having a tip region at which aqueous solution in the channel forms a meniscus, such as the concave meniscus illustrated at 20 in Fig. 2A. The advantages of the open channel construction of the dispenser are discussed below.

With continued reference to Fig. 1, the dispenser device also includes structure for moving the dispenser rapidly toward and away from a support surface, for effecting deposition of a known amount of solution in the dispenser on a support, as will be described below with reference to Figs. 2A-2C. In the embodiment shown, this structure includes a solenoid 22 which is activatable to draw a solenoid piston 24 rapidly downwardly, then release the piston, *e.g.*, under spring bias, to a normal, raised position, as shown. The

dispenser is carried on the piston by a connecting member 26, as shown. The just-described moving structure is also referred to herein as dispensing means for moving the dispenser into engagement with a solid support, for dispensing a known volume of fluid on the support.

The dispensing device just described is carried on an arm 28 that may be moved either linearly or in an x-y plane to position the dispenser at a selected deposition position, as will be described.

Figs. 2A-2C illustrate the method of depositing a known amount of reagent solution in the just-described dispenser on the surface of a solid support, such as the support indicated at 30. The support is a polymer, glass, or other solid-material support having a surface indicated at 31.

In one general embodiment, the surface is a relatively hydrophilic, *i.e.*, wettable surface, such as a surface having native, bound or covalently attached charged groups. On such surface described below is a glass surface having an absorbed layer of a polycationic polymer, such as poly-l-lysine.

In another embodiment, the surface has or is formed to have a relatively hydrophobic character, *i.e.*, one that causes aqueous medium deposited on the surface to bead. A variety of known hydrophobic polymers, such as polystyrene, polypropylene, or polyethylene have desired hydrophobic properties, as do glass and a variety of lubricant or other hydrophobic films that may be applied to the support surface.

Initially, the dispenser is loaded with a selected analyte-specific reagent solution, such as by dipping the dispenser tip, after washing, into a solution of the reagent, and allowing filling by capillary flow into the dispenser channel. The dispenser is now moved

to a selected position with respect to a support surface, placing the dispenser tip directly above the support-surface position at which the reagent is to be deposited. This movement takes place with the dispenser tip in its raised position, as seen in Fig. 2A, where the tip is typically at least several 1-5 mm above the surface of the substrate.

With the dispenser so positioned, solenoid 22 is now activated to cause the dispenser tip to move rapidly toward and away from the substrate surface, making momentary contact with the surface, in effect, tapping the tip of the dispenser against the support surface. The tapping movement of the tip against the surface acts to break the liquid meniscus in the tip channel, bringing the liquid in the tip into contact with the support surface. This, in turn, produces a flowing of the liquid into the capillary space between the tip and the surface, acting to draw liquid out of the dispenser channel, as seen in Fig. 2B.

Fig. 2C shows flow of fluid from the tip onto the support surface, which in this case is a hydrophobic surface. The figure illustrates that liquid continues to flow from the dispenser onto the support surface until it forms a liquid bead 32. At a given bead size, *i.e.*, volume, the tendency of liquid to flow onto the surface will be balanced by the hydrophobic surface interaction of the bead with the support surface, which acts to limit the total bead area on the surface, and by the surface tension of the droplet, which tends toward a given bead curvature. At this point, a given bead volume will have formed, and continued contact of the dispenser tip with the bead, as the dispenser tip is being withdrawn, will have little or no effect on bead volume.

For liquid-dispensing on a more hydrophilic surface, the liquid will have less of a tendency to bead, and the dispensed volume will be more sensitive to the total dwell time of the dispenser tip in the immediate vicinity of the support surface, *e.g.*, the positions illustrated in Figs. 2B and 2C.

The desired deposition volume, *i.e.*, bead volume, formed by this method is preferably in the range 2 pl (picoliters) to 2 nl (nanoliters), although volumes as high as 100 nl or more may be dispensed. It will be appreciated that the selected dispensed volume will depend on (i) the "footprint" of the dispenser tip, *i.e.*, the size of the area spanned by the tip, (ii) the hydrophobicity of the support surface, and (iii) the time of contact with and rate of withdrawal of the tip from the support surface. In addition, bead size may be reduced by increasing the viscosity of the medium, effectively reducing the flow time of liquid from the dispenser onto the support surface. The drop size may be further constrained by depositing the drop in a hydrophilic region surrounded by a hydrophobic grid pattern on the support surface.

In a typical embodiment, the dispenser tip is tapped rapidly against the support surface, with a total residence time in contact with the support of less than about 1 msec, and a rate of upward travel from the surface of about 10 cm/sec.

Assuming that the bead that forms on contact with the surface is a hemispherical bead, with a diameter approximately equal to the width of the dispenser tip, as shown in Fig. 2C, the volume of the bead formed in relation to dispenser tip width (d) is given in Table 1 below. As seen, the volume of the bead ranges between 2 pl to 2 nl as the width size is increased from about 20 to 200 μm.

Table 1

d	Volume (nl)
20 μm	2×10^{-3}
50 μm	3.1×10^{-2}
100 μm	2.5×10^{-1}
200 μm	2

At a given tip size, bead volume can be reduced in a controlled fashion by increasing surface hydrophobicity, reducing time of contact of the tip with the surface, increasing rate of movement of the tip away from the surface, and/or increasing the viscosity of the medium. Once these parameters are fixed, a selected deposition volume in the desired pl to nl range can be achieved in a repeatable fashion.

After depositing a bead at one selected location on a support, the tip is typically moved to a corresponding position on a second support, a droplet is deposited at that position, and this process is repeated until a liquid droplet of the reagent has been deposited at a selected position on each of a plurality of supports.

The tip is then washed to remove the reagent liquid, filled with another reagent liquid and this reagent is now deposited at each another array position on each of the supports. In one embodiment, the tip is washed and refilled by the steps of (i) dipping the capillary channel of the device in a wash solution, (ii) removing wash solution drawn into the capillary channel, and (iii) dipping the capillary channel into the new reagent solution.

From the foregoing, it will be appreciated that the tweezers-like, open-capillary dispenser tip

provides the advantages that (i) the open channel of the tip facilitates rapid, efficient washing and drying before reloading the tip with a new reagent, (ii) passive capillary action can load the sample directly from a standard microwell plate while retaining sufficient sample in the open capillary reservoir for the printing of numerous arrays, (iii) open capillaries are less prone to clogging than closed capillaries, and (iv) open capillaries do not require a perfectly faced bottom surface for fluid delivery.

A portion of a microarray 36 formed on the surface 38 of a solid support 40 in accordance with the method just described is shown in Fig. 3. The array is formed of a plurality of analyte-specific reagent regions, such as regions 42, where each region may include a different analyte-specific reagent. As indicated above, the diameter of each region is preferably between about 20-200 μm. The spacing between each region and its closest (non-diagonal) neighbor, measured from center-to-center (indicated at 44), is preferably in the range of about 20-400 μm. Thus, for example, an array having a center-to-center spacing of about 250 μm contains about 40 regions/cm or 1,600 regions/cm^2. After formation of the array, the support is treated to evaporate the liquid of the droplet forming each region, to leave a desired array of dried, relatively flat regions. This drying may be done by heating or under vacuum.

In some cases, it is desired to first rehydrate the droplets containing the analyte reagents to allow for more time for adsorption to the solid support. It is also possible to spot out the analyte reagents in a humid environment so that droplets do not dry until the arraying operation is complete.

III. Automated Apparatus for Forming Arrays

In another aspect, the invention includes an automated apparatus for forming an array of analyte-assay regions on a solid support, where each region in the array has a known amount of a selected, analyte-specific reagent.

The apparatus is shown in planar, and partially schematic view in Fig. 4. A dispenser device 72 in the apparatus has the basic construction described above with respect to Fig. 1, and includes a dispenser 74 having an open-capillary channel terminating at a tip, substantially as shown in Figs. 1 and 2A-2C.

The dispenser is mounted in the device for movement toward and away from a dispensing position at which the tip of the dispenser taps a support surface, to dispense a selected volume of reagent solution, as described above. This movement is effected by a solenoid 76 as described above. Solenoid 76 is under the control of a control unit 77 whose operation will be described below. The solenoid is also referred to herein as dispensing means for moving the device into tapping engagement with a support, when the device is positioned at a defined array position with respect to that support.

The dispenser device is carried on an arm 74 which is threadedly mounted on a worm screw 80 driven (rotated) in a desired direction by a stepper motor 82 also under the control of unit 77. At its left end in the figure screw 80 is carried in a sleeve 84 for rotation about the screw axis. At its other end, the screw is mounted to the drive shaft of the stepper motor, which in turn is carried on a sleeve 86. The dispenser device, worm screw, the two sleeves mounting the worm screw, and the stepper motor used in moving the device in the "x" (horizontal) direction in the

figure form what is referred to here collectively as a displacement assembly 86.

The displacement assembly is constructed to produce precise, micro-range movement in the direction of the screw, *i.e.*, along an x axis in the figure. In one mode, the assembly functions to move the dispenser in x-axis increments having a selected distance in the range 5-25 μm. In another mode, the dispenser unit may be moved in precise x-axis increments of several microns or more, for positioning the dispenser at associated positions on adjacent supports, as will be described below.

The displacement assembly, in turn, is mounted for movement in the "y" (vertical) axis of the figure, for positioning the dispenser at a selected y axis position. The structure mounting the assembly includes a fixed rod 88 mounted rigidly between a pair of frame bars 90, 92, and a worm screw 94 mounted for rotation between a pair of frame bars 96, 98. The worm screw is driven (rotated) by a stepper motor 100 which operates under the control of unit 77. The motor is mounted on bar 96, as shown.

The structure just described, including worm screw 94 and motor 100, is constructed to produce precise, micro-range movement in the direction of the screw, *i.e.*, along an y axis in the figure. As above, the structure functions in one mode to move the dispenser in y-axis increments having a selected distance in the range 5-250 μm, and in a second mode, to move the dispenser in precise y-axis increments of several microns (μm) or more, for positioning the dispenser at associated positions on adjacent supports.

The displacement assembly and structure for moving this assembly in the y axis are referred to herein collectively as positioning means for positioning the

dispensing device at a selected array position with respect to a support.

A holder 102 in the apparatus functions to hold a plurality of supports, such as supports 104 on which the microarrays of regent regions are to be formed by the apparatus. The holder provides a number of recessed slots, such as slot 106, which receive the supports, and position them at precise selected positions with respect to the frame bars on which the dispenser moving means is mounted.

As noted above, the control unit in the device functions to actuate the two stepper motors and dispenser solenoid in a sequence designed for automated operation of the apparatus in forming a selected microarray of reagent regions on each of a plurality of supports.

The control unit is constructed, according to conventional microprocessor control principles, to provide appropriate signals to each of the solenoid and each of the stepper motors, in a given timed sequence and for appropriate signalling time. The construction of the unit, and the settings that are selected by the user to achieve a desired array pattern, will be understood from the following description of a typical apparatus operation.

Initially, one or more supports are placed in one or more slots in the holder. The dispenser is then moved to a position directly above a well (not shown) containing a solution of the first reagent to be dispensed on the support(s). The dispenser solenoid is actuated now to lower the dispenser tip into this well, causing the capillary channel in the dispenser to fill. Motors 82, 100 are now actuated to position the dispenser at a selected array position at the first of the supports. Solenoid actuation of the dispenser is

then effective to dispense a selected-volume droplet of that reagent at this location. As noted above, this operation is effective to dispense a selected volume preferably between 2 pl and 2 nl of the reagent solution.

The dispenser is now moved to the corresponding position at an adjacent support and a similar volume of the solution is dispensed at this position. The process is repeated until the reagent has been dispensed at this preselected corresponding position on each of the supports.

Where it is desired to dispense a single reagent at more than two array positions on a support, the dispenser may be moved to different array positions at each support, before moving the dispenser to a new support, or solution can be dispensed at individual positions on each support, at one selected position, then the cycle repeated for each new array position.

To dispense the next reagent, the dispenser is positioned over a wash solution (not shown), and the dispenser tip is dipped in and out of this solution until the reagent solution has been substantially washed from the tip. Solution can be removed from the tip, after each dipping, by vacuum, compressed air spray, sponge, or the like.

The dispenser tip is now dipped in a second reagent well, and the filled tip is moved to a second selected array position in the first support. The process of dispensing reagent at each of the corresponding second-array positions is then carried as above. This process is repeated until an entire microarray of reagent solutions on each of the supports has been formed.

IV. Microarray Substrate

This section describes embodiments of a substrate having a microarray of biological polymers carried on the substrate surface. Subsection A describes a multi-cell substrate, each cell of which contains a microarray, and preferably an identical microarray, of distinct biopolymers, such as distinct polynucleotides, formed on a porous surface. Subsection B describes a microarray of distinct polynucleotides bound on a glass slide coated with a polycationic polymer.

A. Multi-Cell Substrate

Fig. 9 illustrates, in plan view, a substrate 110 constructed according to the invention. The substrate has an 8 x 12 rectangular array 112 of cells, such as cells 114, 116, formed on the substrate surface. With reference to Fig. 10, each cell, such as cell 114, in turn supports a microarray 118 of distinct biopolymers, such as polypeptides or polynucleotides at known, addressable regions of the microarray. Two such regions forming the microarray are indicated at 120, and correspond to regions, such as regions 42, forming the microarray of distinct biopolymers shown in Fig. 3.

The 96-cell array shown in Fig. 9 has typically array dimensions between about 12 and 244 mm in width and 8 and 400 mm in length, with the cells in the array having width and length dimension of 1/12 and 1/8 the array width and length dimensions, respectively, *i.e.*, between about 1 and 20 in width and 1 and 50 mm in length.

The construction of substrate is shown cross-sectionally in Fig. 11, which is an enlarged sectional view taken along view line 124 in Fig. 9. The substrate includes a water-impermeable backing 126, such as a glass slide or rigid polymer sheet. Formed on the surface of the backing is a water-permeable film

128. The film is formed of a porous membrane material, such as nitrocellulose membrane, or a porous web material, such as a nylon, polypropylene, or PVDF porous polymer material. The thickness of the film is preferably between about 10 and 1000 μm. The film may be applied to the backing by spraying or coating uncured material on the backing, or by applying a preformed membrane to the backing. The backing and film may be obtained as a preformed unit from commercial source, e.g., a plastic-backed nitrocellulose film available from Schleicher and Schuell Corporation.

With continued reference to Fig. 11, the film-covered surface in the substrate is partitioned into a desired array of cells by water-impermeable grid lines, such as lines 130, 132, which have infiltrated the film down to the level of the backing, and extend above the surface of the film as shown, typically a distance of 100 to 2000 μm above the film surface.

The grid lines are formed on the substrate by laying down an uncured or otherwise flowable resin or elastomer solution in an array grid, allowing the material to infiltrate the porous film down to the backing, then curing or otherwise hardening the grid lines to form the cell-array substrate.

One preferred material for the grid is a flowable silicone available from Loctite Corporation. The barrier material can be extruded through a narrow syringe (e.g., 22 gauge) using air pressure or mechanical pressure. The syringe is moved relative to the solid support to print the barrier elements as a grid pattern. The extruded bead of silicone wicks into the pores of the solid support and cures to form a shallow waterproof barrier separating the regions of the solid support.

In alternative embodiments, the barrier element can be a wax-based material or a thermoset material such as epoxy. The barrier material can also be a UV-curing polymer which is exposed to UV light after being printed onto the solid support. The barrier material may also be applied to the solid support using printing techniques such as silk-screen printing. The barrier material may also be a heat-seal stamping of the porous solid support which seals its pores and forms a water-impervious barrier element. The barrier material may also be a shallow grid which is laminated or otherwise adhered to the solid support.

In addition to plastic-backed nitrocellulose, the solid support can be virtually any porous membrane with or without a non-porous backing. Such membranes are readily available from numerous vendors and are made from nylon, PVDF, polysulfone and the like. In an alternative embodiment, the barrier element may also be used to adhere the porous membrane to a non-porous backing in addition to functioning as a barrier to prevent cross contamination of the assay reagents.

In an alternative embodiment, the solid support can be of a non-porous material. The barrier can be printed either before or after the microarray of biomolecules is printed on the solid support.

As can be appreciated, the cells formed by the grid lines and the underlying backing are water-impermeable, having side barriers projecting above the porous film in the cells. Thus, defined-volume samples can be placed in each well without risk of cross-contamination with sample material in adjacent cells. In Fig. 11, defined volumes samples, such as sample 134, are shown in the cells.

As noted above, each well contains a microarray of distinct biopolymers. In one general embodiment, the

microarrays in the well are identical arrays of distinct biopolymers, *e.g.*, different sequence polynucleotides. Such arrays can be formed in accordance with the methods described in Section II, by depositing a first selected polynucleotide at the same selected microarray position in each of the cells, then depositing a second polynucleotide at a different microarray position in each well, and so on until a complete, identical microarray is formed in each cell.

In a preferred embodiment, each microarray contains about 10^3 distinct polynucleotide or polypeptide biopolymers per surface area of less than about 1 cm^2. Also in a preferred embodiment, the biopolymers in each microarray region are present in a defined amount between about 0.1 femtomoles and 100 nanomoles. The ability to form high-density arrays of biopolymers, where each region is formed of a well-defined amount of deposited material, can be achieved in accordance with the microarray-forming method described in Section II.

Also in a preferred embodiments, the biopolymers are polynucleotides having lengths of at least about 50 bp, *i.e.*, substantially longer than oligonucleotides which can be formed in high-density arrays by schemes involving parallel, step-wise polymer synthesis on the array surface.

In the case of a polynucleotide array, in an assay procedure, a small volume of the labeled DNA probe mixture in a standard hybridization solution is loaded onto each cell. The solution will spread to cover the entire microarray and stop at the barrier elements. The solid support is then incubated in a humid chamber at the appropriate temperature as required by the assay.

Each assay may be conducted in an "open-face" format where no further sealing step is required, since the hybridization solution will be kept properly hydrated by the water vapor in the humid chamber. At the conclusion of the incubation step, the entire solid support containing the numerous microarrays is rinsed quickly enough to dilute the assay reagents so that no significant cross contamination occurs. The entire solid support is then reacted with detection reagents if needed and analyzed using standard colorimetric, radioactive or fluorescent detection means. All processing and detection steps are performed simultaneously to all of the microarrays on the solid support ensuring uniform assay conditions for all of the microarrays on the solid support.

B. Glass-Slide Polynucleotide Array

Fig. 5 shows a substrate 136 formed according to another aspect of the invention, and intended for use in detecting binding of labeled polynucleotides to one or more of a plurality distinct polynucleotides. The substrate includes a glass substrate 138 having formed on its surface, a coating of a polycationic polymer, preferably a cationic polypeptide, such as polylysine or polyarginine. Formed on the polycationic coating is a microarray 140 of distinct polynucleotides, each localized at known selected array regions, such as regions 142.

The slide is coated by placing a uniform-thickness film of a polycationic polymer, *e.g.*, poly-l-lysine, on the surface of a slide and drying the film to form a dried coating. The amount of polycationic polymer added is sufficient to form at least a monolayer of polymers on the glass surface. The polymer film is bound to surface via electrostatic binding between

negative silyl-OH groups on the surface and charged amine groups in the polymers. Poly-l-lysine coated glass slides may be obtained commercially, *e.g.*, from Sigma Chemical Co. (St. Louis, MO).

To form the microarray, defined volumes of distinct polynucleotides are deposited on the polymer-coated slide, as described in Section II. According to an important feature of the substrate, the deposited polynucleotides remain bound to the coated slide surface non-covalently when an aqueous DNA sample is applied to the substrate under conditions which allow hybridization of reporter-labeled polynucleotides in the sample to complementary-sequence (single-stranded) polynucleotides in the substrate array. The method is illustrated in Examples 1 and 2.

To illustrate this feature, a substrate of the type just described, but having an array of same-sequence polynucleotides, was mixed with fluorescent-labeled complementary DNA under hybridization conditions. After washing to remove non-hybridized material, the substrate was examined by low-power fluorescence microscopy. The array can be visualized by the relatively uniform labeling pattern of the array regions.

In a preferred embodiment, each microarray contains at least 10^3 distinct polynucleotide or polypeptide biopolymers per surface area of less than about 1 cm^2. In the embodiment shown in Fig. 5, the microarray contains 400 regions in an area of about 16 mm^2, or 2.5×10^3 regions/cm^2. Also in a preferred embodiment, the polynucleotides in the each microarray region are present in a defined amount between about 0.1 femtomoles and 100 nanomoles in the case of polynucleotides. As above, the ability to form high-

density arrays of this type, where each region is formed of a well-defined amount of deposited material, can be achieved in accordance with the microarray-forming method described in Section II.

Also in a preferred embodiments, the polynucleotides have lengths of at least about 50 bp, *i.e.*, substantially longer than oligonucleotides which can be formed in high-density arrays by various in situ synthesis schemes.

V. Utility

Microarrays of immobilized nucleic acid sequences prepared in accordance with the invention can be used for large scale hybridization assays in numerous genetic applications, including genetic and physical mapping of genomes, monitoring of gene expression, DNA sequencing, genetic diagnosis, genotyping of organisms, and distribution of DNA reagents to researchers.

For gene mapping, a gene or a cloned DNA fragment is hybridized to an ordered array of DNA fragments, and the identity of the DNA elements applied to the array is unambiguously established by the pixel or pattern of pixels of the array that are detected. One application of such arrays for creating a genetic map is described by Nelson, *et al.* (1993). In constructing physical maps of the genome, arrays of immobilized cloned DNA fragments are hybridized with other cloned DNA fragments to establish whether the cloned fragments in the probe mixture overlap and are therefore contiguous to the immobilized clones on the array. For example, Lehrach, *et al.*, describe such a process.

The arrays of immobilized DNA fragments may also be used for genetic diagnostics. To illustrate, an array containing multiple forms of a mutated gene or genes can be probed with a labeled mixture of a

patient's DNA which will preferentially interact with only one of the immobilized versions of the gene.

The detection of this interaction can lead to a medical diagnosis. Arrays of immobilized DNA fragments can also be used in DNA probe diagnostics. For example, the identity of a pathogenic microorganism can be established unambiguously by hybridizing a sample of the unknown pathogen's DNA to an array containing many types of known pathogenic DNA. A similar technique can also be used for unambiguous genotyping of any organism. Other molecules of genetic interest, such as cDNA's and RNA's can be immobilized on the array or alternately used as the labeled probe mixture that is applied to the array.

In one application, an array of cDNA clones representing genes is hybridized with total cDNA from an organism to monitor gene expression for research or diagnostic purposes. Labeling total cDNA from a normal cell with one color fluorophore and total cDNA from a diseased cell with another color fluorophore and simultaneously hybridizing the two cDNA samples to the same array of cDNA clones allows for differential gene expression to be measured as the ratio of the two fluorophore intensities. This two-color experiment can be used to monitor gene expression in different tissue types, disease states, response to drugs, or response to environmental factors. & An example of this approach is illustrated in Examples 2, described with respect to Fig. 8.

By way of example and without implying a limitation of scope, such a procedure could be used to simultaneously screen many patients against all known mutations in a disease gene. This invention could be used in the form of, for example, 96 identical 0.9 cm × 2.2 cm microarrays fabricated on a single 12 cm × 18 cm

sheet of plastic-backed nitrocellulose where each microarray could contain, for example, 100 DNA fragments representing all known mutations of a given gene. The region of interest from each of the DNA samples from 96 patients could be amplified, labeled, and hybridized to the 96 individual arrays with each assay performed in 100 microliters of hybridization solution. The approximately 1 thick silicone rubber barrier elements between individual arrays prevent cross contamination of the patient samples by sealing the pores of the nitrocellulose and by acting as a physical barrier between each microarray. The solid support containing all 96 microarrays assayed with the 96 patient samples is incubated, rinsed, detected and analyzed as a single sheet of material using standard radioactive, fluorescent, or colorimetric detection means (Maniatas, *et al.*, 1989). Previously, such a procedure would involve the handling, processing and tracking of 96 separate membranes in 96 separate sealed chambers. By processing all 96 arrays as a single sheet of material, significant time and cost savings are possible.

The assay format can be reversed where the patient or organism's DNA is immobilized as the array elements and each array is hybridized with a different mutated allele or genetic marker. The gridded solid support can also be used for parallel non-DNA ELISA assays. Furthermore, the invention allows for the use of all standard detection methods without the need to remove the shallow barrier elements to carry out the detection step.

In addition to the genetic applications listed above, arrays of whole cells, peptides, enzymes, antibodies, antigens, receptors, ligands, phospholipids, polymers, drug cogener preparations or

chemical substances can be fabricated by the means described in this invention for large scale screening assays in medical diagnostics, drug discovery, molecular biology, immunology and toxicology.

The multi-cell substrate aspect of the invention allows for the rapid and convenient screening of many DNA probes against many ordered arrays of DNA fragments. This eliminates the need to handle and detect many individual arrays for performing mass screenings for genetic research and diagnostic applications. Numerous microarrays can be fabricated on the same solid support and each microarray reacted with a different DNA probe while the solid support is processed as a single sheet of material.

The following examples illustrate, but in no way are intended to limit, the present invention.

Example 1

Genomic-Complexity Hybridization to Micro DNA Arrays Representing the Yeast *Saccharomyces cerevisiae* Genome with Two-Color Fluorescent Detection

The array elements were randomly amplified PCR (Bohlander, *et al.*, 1992) products using physically mapped lambda clones of *S. cerevisiae* genomic DNA templates (Riles, *et al.*, 1993). The PCR was performed directly on the lambda phage lysates resulting in an amplification of both the 35 kb lambda vector and the 5-15 kb yeast insert sequences in the form of a uniform distribution of PCR product between 250-1500 base pairs in length. The PCR product was purified using Sephadex G50 gel filtration (Pharmacia, Piscataway, NJ) and concentrated by evaporation to dryness at room temperature overnight. Each of the 864 amplified

lambda clones was rehydrated in 15 μl of 3 × SSC in preparation for spotting onto the glass.

The micro arrays were fabricated on microscope slides which were coated with a layer of poly-l-lysine (Sigma). The automated apparatus described in Section IV loaded 1 μl of the concentrated lambda clone PCR product in 3 × SSC directly from 96 well storage plates into the open capillary printing element and deposited ~5 nl of sample per slide at 380 micron spacing between spots, on each of 40 slides. The process was repeated for all 864 samples and 8 control spots. After the spotting operation was complete, the slides were rehydrated in a humid chamber for 2 hours, baked in a dry 80° vacuum oven for 2 hours, rinsed to remove un-absorbed DNA and then treated with succinic anhydride to reduce non-specific adsorption of the labeled hybridization probe to the poly-l-lysine coated glass surface. Immediately prior to use, the immobilized DNA on the array was denatured in distilled water at 90° for 2 minutes.

For the pooled chromosome experiment, the 16 chromosomes of *Saccharomyces cerevisiae* were separated in a CHEF agarose gel apparatus (Biorad, Richmond, CA). The six largest chromosomes were isolated in one gel slice and the smallest 10 chromosomes in a second gel slice. The DNA was recovered using a gel extraction kit (Qiagen, Chatsworth, CA). The two chromosome pools were randomly amplified in a manner similar to that used for the target lambda clones. Following amplification, 5 micrograms of each of the amplified chromosome pools were separately random-primer labeled using Klenow polymerase (Amersham, Arlington Heights, IL) with a lissamine conjugated nucleotide analog (Dupont NEN, Boston, MA) for the pool containing the six largest chromosomes, and with a fluorescein

conjugated nucleotide analog (BMB) for the pool containing smallest ten chromosomes. The two pools were mixed and concentrated using an ultrafiltration device (Amicon, Danvers, MA).

Five micrograms of the hybridization probe consisting of both chromosome pools in 7.5 μl of TE was denatured in a boiling water bath and then snap cooled on ice. 2.5 μl of concentrated hybridization solution (5 × SSC and 0.1% SDS) was added and all 10 μl transferred to the array surface, covered with a cover slip, placed in a custom-built single-slide humidity chamber and incubated at 60° for 12 hours. The slides were then rinsed at room temperature in 0.1 × SSC and 0.1%SDS for 5 minutes, cover slipped and scanned.

A custom built laser fluorescent scanner was used to detect the two-color hybridization signals from the 1.8 × 1.8 cm array at 20 micron resolution. The scanned image was gridded and analyzed using custom image analysis software. After correcting for optical crosstalk between the fluorophores due to their overlapping emission spectra, the red and green hybridization values for each clone on the array were correlated to the known physical map position of the clone resulting in a computer-generated color karyotype of the yeast genome.

Figure 6 shows the hybridization pattern of the two chromosome pools. A red signal indicates that the lambda clone on the array surface contains a cloned genomic DNA segment from one of the largest six yeast chromosomes. A green signal indicates that the lambda clone insert comes from one of the smallest ten yeast chromosomes. Orange signals indicate repetitive sequences which cross hybridized to both chromosome pools. Control spots on the array confirm that the hybridization is specific and reproducible.

The physical map locations of the genomic DNA fragments contained in each of the clones used as array elements have been previously determined by Olson and co-workers (Riles, *et al.*) allowing for the automatic generation of the color karyotype shown in Figure 7. The color of a chromosomal section on the karyotype corresponds to the color of the array element containing the clone from that section. The black regions of the karyotype represent false negative dark spots on the array (10%) or regions of the genome not covered by the Olson clone library (90%). Note that the largest six chromosomes are mainly red while the smallest ten chromosomes are mainly green matching the original CHEF gel isolation of the hybridization probe. Areas of the red chromosomes containing green spots and vice-versa are probably due to spurious sample tracking errors in the formation of the original library and in the amplification and spotting procedures.

The yeast genome arrays have also been probed with individual clones or pools of clones that are fluorescently labeled for physical mapping purposes. The hybridization signals of these clones to the array were translated into a position on the physical map of yeast.

Example 2

Total cDNA Hybridized to Micro Arrays of cDNA Clones with Two-Color Fluorescent Detection

24 clones containing cDNA inserts from the plant *Arabidopsis* were amplified using PCR. Salt was added to the purified PCR products to a final concentration of 3 × SSC. The cDNA clones were spotted on poly-l-lysine coated microscope slides in a manner similar to Example 1. Among the cDNA clones was a clone

representing a transcription factor HAT 4, which had previously been used to create a transgenic line of the plant *Arabidopsis*, in which this gene is present at ten times the level found in wild-type *Arabidopsis* (Schena, *et al.*, 1992).

Total poly-A mRNA from wild type *Arabidopsis* was isolated using standard methods (Maniatis, *et al.*, 1989) and reverse transcribed into total cDNA, using fluorescein nucleotide analog to label the cDNA product (green fluorescence). A similar procedure was performed with the transgenic line of *Arabidopsis* where the transcription factor HAT4 was inserted into the genome using standard gene transfer protocols. cDNA copies of mRNA from the transgenic plant are labeled with a lissamine nucleotide analog (red fluorescence). Two micrograms of the cDNA products from each type of plant were pooled together and hybridized to the cDNA clone array in a 10 microliter hybridization reaction in a manner similar to Example 1. Rinsing and detection of hybridization was also performed in a manner similar to Example 1. Fig. 8 show the resulting hybridization pattern of the array.

Genes equally expressed in wild type and the transgenic *Arabidopsis* appeared yellow due to equal contributions of the green and red fluorescence to the final signal. The dots are different intensities of yellow indicating various levels of gene expression. The cDNA clone representing the transcription factor HAT4, expressed in the transgenic line of *Arabidopsis* but not detectably expressed in wild type *Arabidopsis*, appears as a red dot (with the arrow pointing to it), indicating the preferential expression of the transcription factor in the red-labeled transgenic *Arabidopsis* and the relative lack of expression of the

transcription factor in the green-labeled wild type *Arabidopsis*.

An advantage of the microarray hybridization format for gene expression studies is the high partial concentration of each cDNA species achievable in the 10 microliter hybridization reaction. This high partial concentration allows for detection of rare transcripts without the need for PCR amplification of the hybridization probe which may bias the true genetic representation of each discrete cDNA species.

Gene expression studies such as these can be used for genomics research to discover which genes are expressed in which cell types, disease states, development states or environmental conditions. Gene expression studies can also be used for diagnosis of disease by empirically correlating gene expression patterns to disease states.

Example 3

Multiplexed Colorimetric Hybridization on a Gridded Solid Support

A sheet of plastic-backed nitrocellulose was gridded with barrier elements made from silicone rubber according to the description in Section IV-A. The sheet was soaked in 10 × SSC and allowed to dry. As shown in Fig. 12, 192 M13 clones each with a different yeast inserts were arrayed 400 microns apart in four quadrants of the solid support using the automated device described in Section III. The bottom left quadrant served as a negative control for hybridization while each of the other three quadrants was hybridized simultaneously with a different oligonucleotide using the open-face hybridization technology described in Section IV-A. The first two and last four elements of

each array are positive controls for the colorimetric detection step.

The oligonucleotides were labeled with fluorescein which was detected using an anti-fluorescein antibody conjugated to alkaline phosphatase that precipitated an NBT/BCIP dye on the solid support (Amersham). Perfect matches between the labeled oligos and the M13 clones resulted in dark spots visible to the naked eye and detected using an optical scanner (HP ScanJet II) attached to a personal computer. The hybridization patterns are different in every quadrant indicating that each oligo found several unique M13 clones from among the 192 with a perfect sequence match. Note that the open capillary printing tip leaves detectable dimples on the nitrocellulose which can be used to automatically align and analyze the images.

Although the invention has been described with respect to specific embodiments and methods, it will be clear that various changes and modification may be made without departing from the invention.

IT IS CLAIMED:

1. A method of forming a microarray of analyte-assay regions on a solid support, where each region in the array has a known amount of a selected, analyte-specific reagent, said method comprising,

(a) loading a solution of a selected analyte-specific reagent in a reagent-dispensing device having an elongate capillary channel (i) formed by spaced-apart, coextensive elongate members, (ii) adapted to hold a quantity of the reagent solution and (iii) having a tip region at which aqueous solution in the channel forms a meniscus,

(b) tapping the tip of the dispensing device against a solid support at a defined position on the surface, with an impulse effective to break the meniscus in the capillary channel and deposit a selected volume of solution on the surface, and

(c) repeating steps (a) and (b) until said array is formed.

2. The method of claim 1, wherein said tapping is carried out with an impulse effective to deposit a selected volume in the volume range between 0.01 to 100 nl.

3. The method of claim 1, wherein said channel is formed by a pair of spaced-apart tapered elements.

4. The method of claim 1, for forming a plurality of such arrays, wherein step (b) is applied to a selected position on each of a plurality of solid supports at each repeat cycle proceeding step (c).

5. The method of claim 1, which further includes, after performing steps (a) and (b) at least one time, reloading the reagent-dispensing device with a new reagent solution by the steps of (i) dipping the capillary channel of the device in a wash solution, (ii) removing wash solution drawn into the capillary channel, and (iii) dipping the capillary channel into the new reagent solution.

6. Automated apparatus for forming a microarray of analyte-assay regions on a plurality of solid supports, where each region in the array has a known amount of a selected, analyte-specific reagent, said apparatus comprising

(a) a holder for holding, at known positions, a plurality of planar supports,

(b) a reagent dispensing device having an open capillary channel (i) formed by spaced-apart, coextensive elongate members (ii) adapted to hold a quantity of the reagent solution and (iii) having a tip region at which aqueous solution in the channel forms a meniscus,

(c) positioning means for positioning the dispensing device at a selected array position with respect to a support in said holder,

(d) dispensing means for moving the device into tapping engagement against a support with a selected impulse, when the device is positioned at a defined array position with respect to that support, with an impulse effective to break the meniscus of liquid in the capillary channel and deposit a selected volume of solution on the surface, and

(e) control means for controlling said positioning and dispensing means.

7. The apparatus of claim 6, wherein said dispensing means is effective to move said dispensing device against a support with an impulse effective to deposit a selected volume in the volume range between 0.01 to 100 nl.

8. The apparatus of claim 6, wherein said channel is formed by a pair of spaced-apart tapered elements.

9. The apparatus of claim 6, wherein the control means operates to (i) place the dispensing device at a loading station, (ii) move the capillary channel in the device into a selected reagent at the loading station, to load the dispensing device with the reagent, and (iii) dispense the reagent at a defined array position on each of the supports on said holder.

10. The apparatus of claim 6, wherein the control device further operates, at the end of a dispensing cycle, to wash the dispensing device by (i) placing the dispensing device at a washing station, (ii) moving the capillary channel in the device into a wash fluid, to load the dispensing device with the fluid, and (iii) remove the wash fluid prior to loading the dispensing device with a fresh selected reagent.

11. The apparatus of claim 6, wherein said device is one of a plurality of such devices which are carried on the arm for dispensing different analyte assay reagents at selected spaced array positions.

12. A substrate with a surface having a microarray of at least 10^3 distinct polynucleotide or polypeptide biopolymers per 1 cm^2 surface area, each

distinct biopolymer sample (i) being disposed at a separate, defined position in said array, (ii) having a length of at least 50 subunits, and (iii) being present in a defined amount between about 0.1 femtomole and 100 nanomoles.

13. The substrate of claim 12, wherein said surface is glass slide coated with polylysine, and said biopolymers are polynucleotides.

14. The substrate of claim 12, wherein said substrate has a water-impermeable backing, a water-permeable film formed on the backing, and a grid formed on the film, where said grid (i) is composed of intersecting water-impervious grid elements extending from said backing to positions raised above the surface of said film, and (ii) partitions the film into a plurality of water-impervious cells, where each cell contains such a biopolymer array.

15. A substrate with a surface array of sample-receiving cells, comprising

a water-impermeable backing,

a water-permeable film formed on the backing, and a grid formed on the film, said grid being composed of intersecting water-impervious grid elements extending from said backing to positions raised above the surface of said film.

16. The substrate of claim 15, wherein the cells of the array each contain an array of biopolymers.

17. A substrate for use in detecting binding of labeled biopolymers to one or more of a plurality distinct polynucleotides, comprising

a non-porous, glass substrate,

a coating of a cationic polymer on said substrate, and

an array of distinct polynucleotides to said coating, where each biopolymer is disposed at a separate, defined position in a surface array of biopolymers.

18. A method of detecting differential expression of each of a plurality of genes in a first cell type with respect to expression of the same genes in a second cell types, said method comprising

producing fluorescence-labeled cDNA's from mRNA's isolated from the two cells types, where the cDNA's from the first and second cells are labeled with first and second different fluorescent reporters,

adding a mixture of the labeled cDNA's from the two cell types to an array of polynucleotides representing a plurality of known genes derived from the two cell types, under conditions that result in hybridization of the cDNA's to complementary-sequence polynucleotides in the array; and

examining the array by fluorescence under fluorescence excitation conditions in which (i) polynucleotides in the array that are hybridized predominantly to cDNA's derived from one of the first and second cell types give a distinct first or second fluorescence emission color, respectively, and (ii) polynucleotides in the array that are hybridized to substantially equal numbers of cDNA's derived from the first and second cell types give a distinct combined fluorescence emission color, respectively,

wherein the relative expression of known genes in the two cell types can be determined by the observed fluorescence emission color of each spot.

19. The method of claim 18, wherein the array of polynucleotides is formed on a substrate with a surface having an array of at least 10^2 distinct polynucleotide or polypeptide biopolymers in a surface area of less than about 1 cm^2, each distinct biopolymer (i) being disposed at a separate, defined position in said array, (ii) having a length of at least 50 subunits, and (iii) being present in a defined amount between about .1 femtomole and 100 nmoles.

20. The method of claim 19, wherein said surface is a glass slide coated with polylysine, and said biopolymers are polynucleotides non-covalently bound to said polylysine.

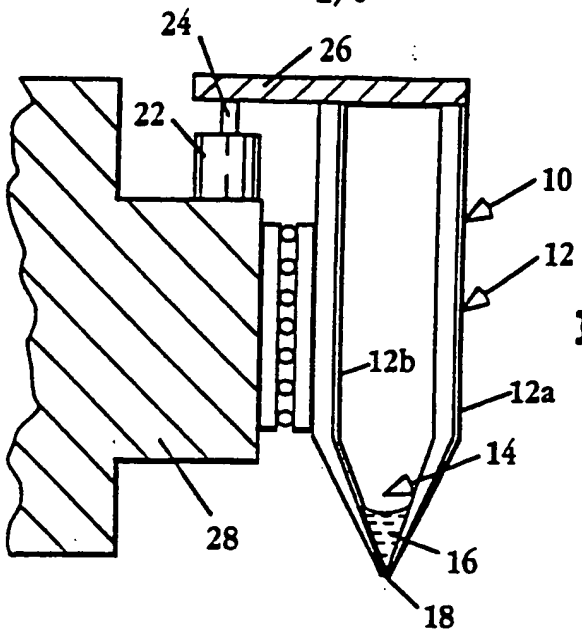


Fig. 1

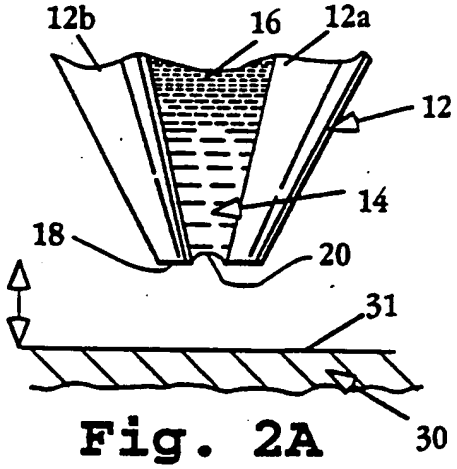


Fig. 2A

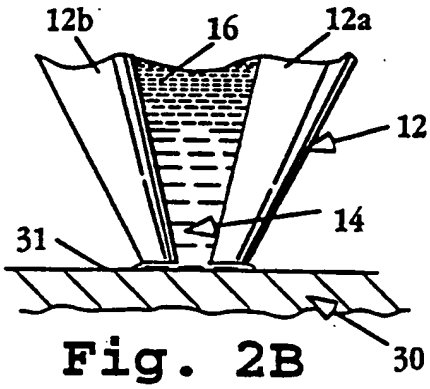


Fig. 2B

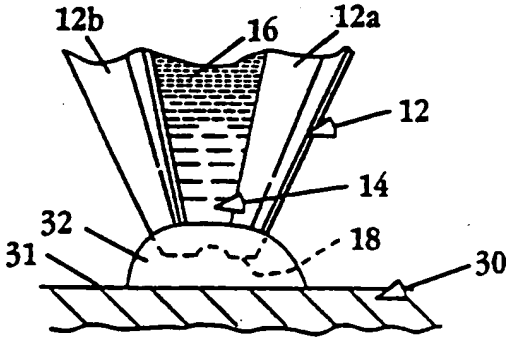


Fig. 2C

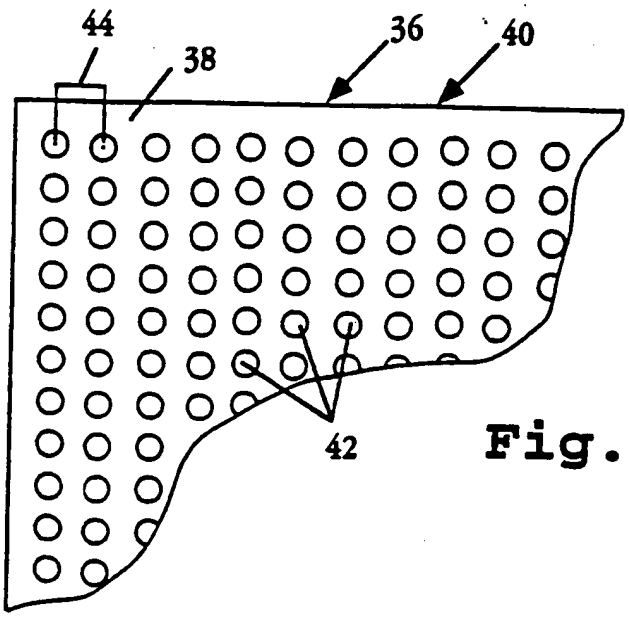

44
38
36
40
42

Fig. 3

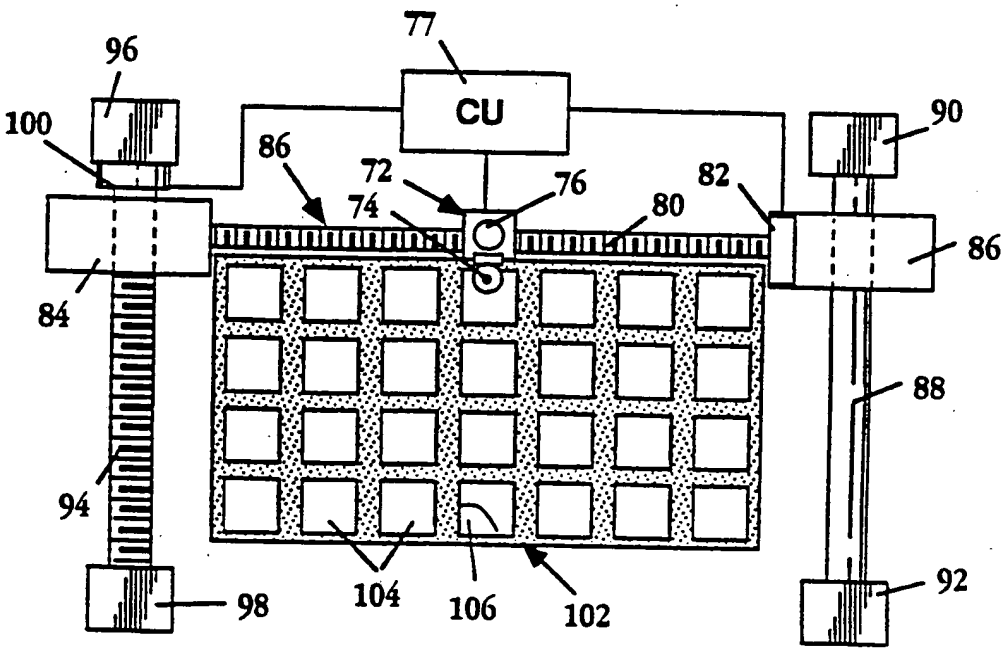

77
CU
96
100
86
72
76
74
80
82
90
86
84
88
94
98
104
106
102
92

Fig. 4

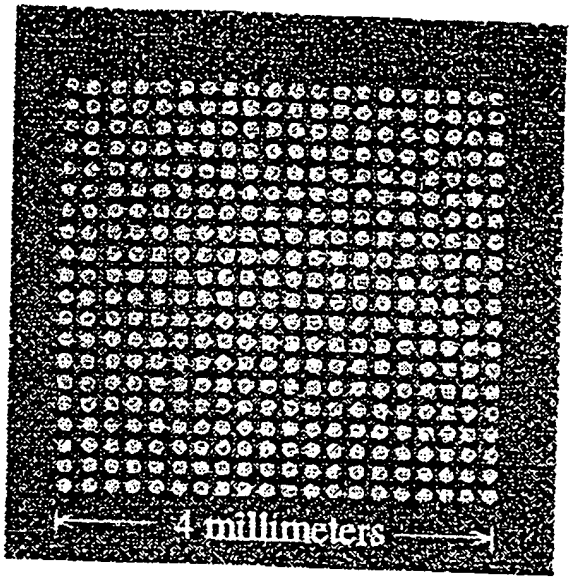


Fig. 5

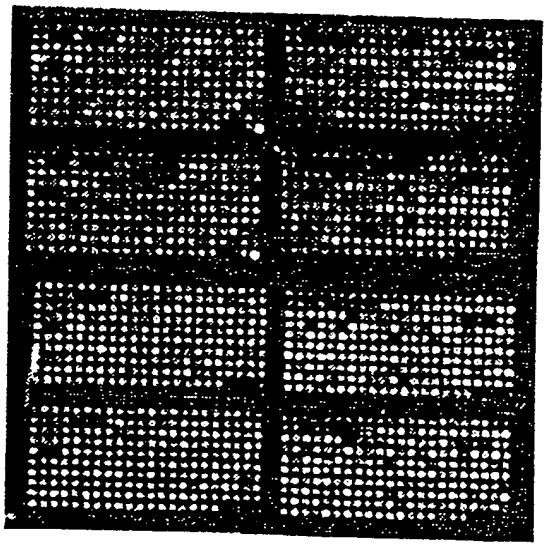


Fig. 6

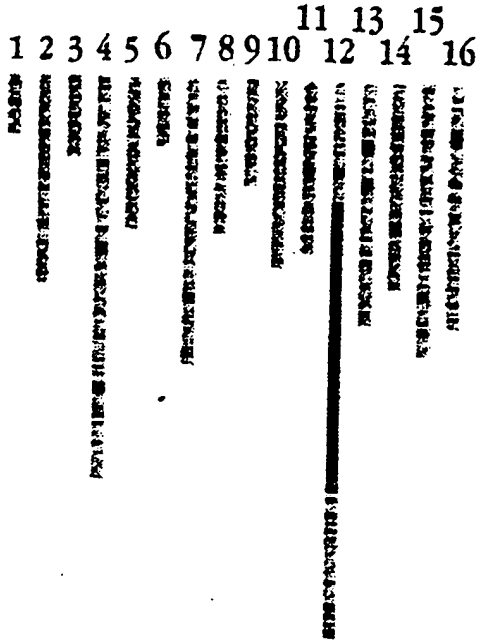


Fig. 7

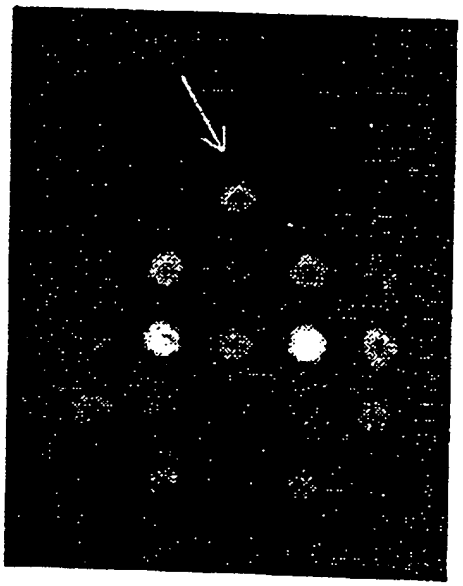


Fig. 8

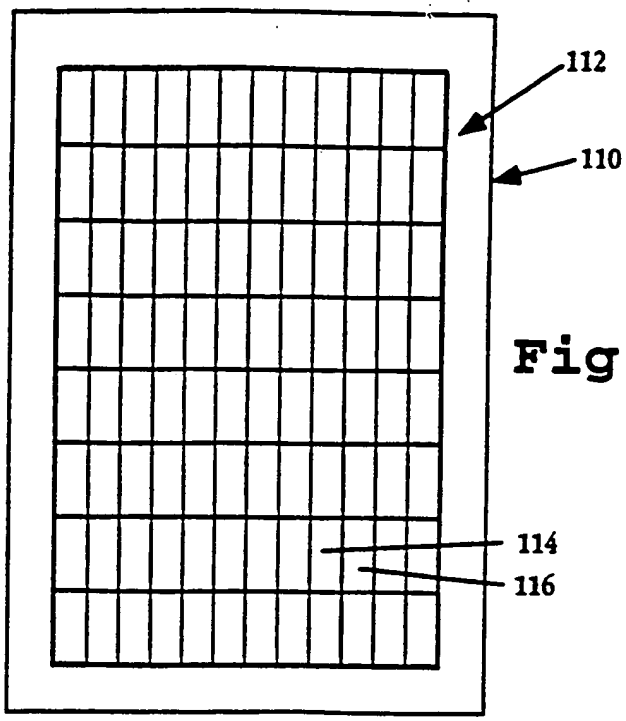


Fig. 9

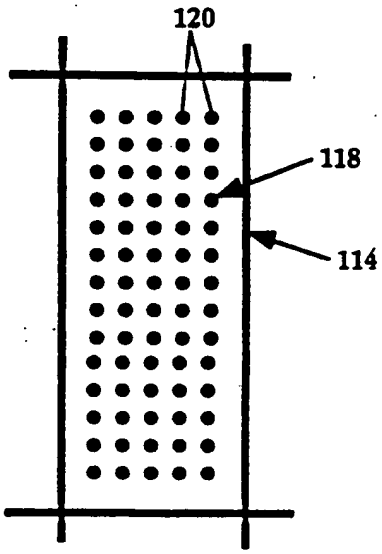


Fig. 10

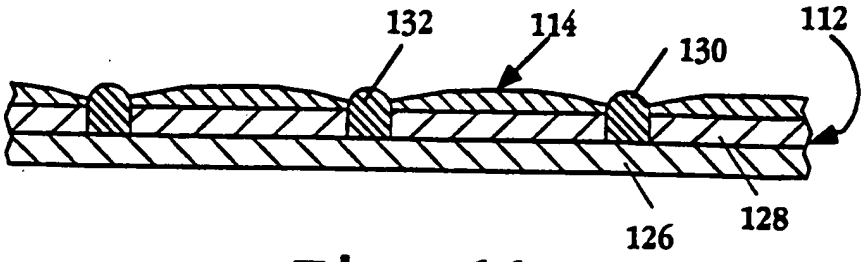


Fig. 11

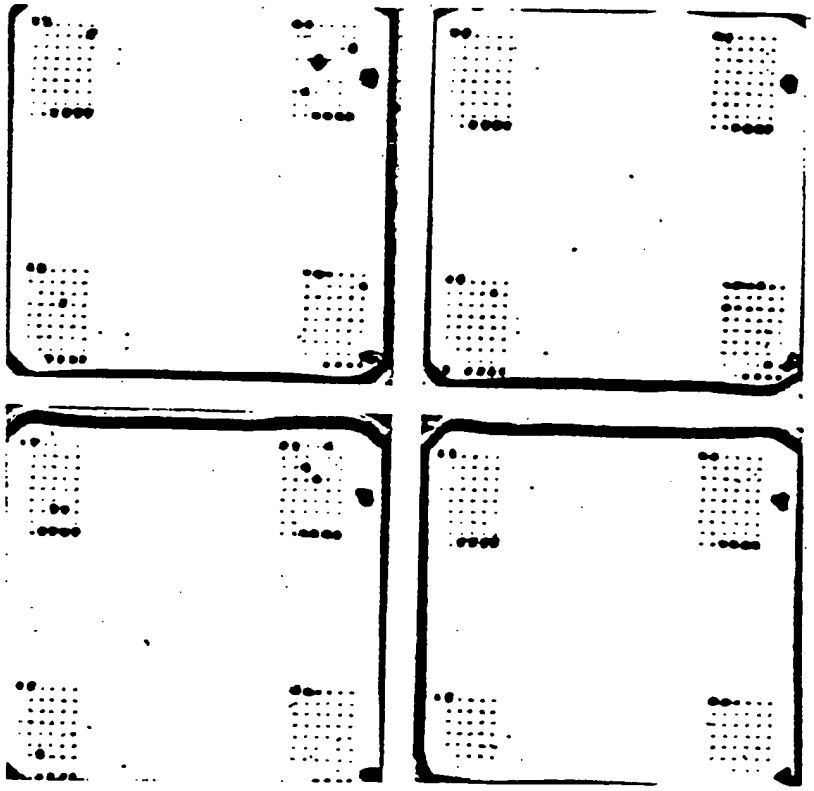


Fig. 12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/07659

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :G01N 33/543, 33/68
US CL :435/6; 436/518

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 422/57; 435/4.6,973; 436/518,524,527,531,805,809

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A,P	US, A, 5,338,688 (DEEG ET AL) 16 August 1994, see entire document	1-17
A	US, A, 5,204,268 (MATSUMOTO) 20 April 1993, see entire document.	6-11
A	US, A, 4,071,315 (CHATEAU) 31 January 1978, see entire document.	12-17
A	US, A, 5,100,777 (CHANG) 31 March 1992, see entire document.	12-17
A	US, A, 5,200,312 (OPRANDY) 06 April 1993, see entire document.	12-17

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

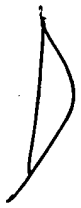
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 SEPTEMBER 1995	06 OCT 1995
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer CHRISTOPHER CHIN Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*



US005807522A

United States Patent [19]

Brown et al.

[11] Patent Number: 5,807,522

[45] Date of Patent: Sep. 15, 1998

[54] METHODS FOR FABRICATING MICROARRAYS OF BIOLOGICAL SAMPLES

[75] Inventors: Patrick O. Brown, Stanford; Tidhar Dari Shalon, Atherton, both of Calif.

[73] Assignee: The Board of Trustees of the Leland Stanford Junior University, Stanford, Calif.

[21] Appl. No.: 477,809

[22] Filed: Jun. 7, 1995

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 261,388, Jun. 17, 1994, abandoned.

[51] Int. Cl.[6] C12M 1/34; C12M 1/40

[52] U.S. Cl. 422/50; 422/52; 422/55; 422/56; 422/57; 422/68.1; 422/69; 422/82.05; 422/82.06; 422/82.07; 422/82.08; 435/6; 435/7.1; 436/501; 530/300; 530/333; 530/334; 530/350; 536/25.3

[58] Field of Search 435/6, 7.1, 172.3; 536/23.1, 24.31, 25.3; 935/78, 3, 19, 80; 436/501, 813; 422/50, 52, 55, 56, 57, 68.1, 69, 82.05, 82.06–82.08; 530/300, 333, 334, 350

[56] References Cited

U.S. PATENT DOCUMENTS

3,730,844	5/1973	Gilham et al.	435/6
4,071,315	1/1978	Chateau	436/518
4,486,539	12/1984	Ranki et al.	436/504
4,556,643	12/1985	Paau et al.	435/5
4,563,419	1/1986	Ranki et al.	435/6
4,591,570	5/1986	Chang	436/518
4,670,380	6/1987	Dattagupta	435/6
4,677,054	6/1987	White et al.	435/6
4,683,195	7/1987	Mullis et al.	435/6
4,683,202	7/1987	Mullis	435/91.2
4,716,106	12/1987	Chiswell	435/6

(List continued on next page.)

FOREIGN PATENT DOCUMENTS

721016A2	7/1996	European Pat. Off. .
WO 90/03382	4/1990	WIPO .
WO 92/10588	6/1992	WIPO .
WO 93/22680	11/1993	WIPO .
WO 95/00530	1/1995	WIPO .
WO 95/15970	6/1995	WIPO .
WO 95/21944	8/1995	WIPO .
WO 95/25116	9/1995	WIPO .
WO 96/17958	6/1996	WIPO .

OTHER PUBLICATIONS

Billings et al., "New Techniques for Physical Mapping of the Human Genome," *FASEB*, 5:28–34 (1991).

Chee, et al., "Accessing Genetic Information with High-Density DNA Arrays", *Science*, 274:610–614 (1996).

Drmanac et al., "DNA Sequence Determination by Hybridization: A Strategy for Efficient Large-Scale Sequencing, "*Science*, 260:1649–1652 (1993).

Drmanac et al., "Laboratory Methods: Reliable Hybridization of Oligonucleotides as Short as Six Nucleotides," *DNA and Cell Biology*, 9:527–534 (1990).

Drmanac et al., "Sequencing by Hybridization: Towards an Automated Sequencing of One Million M13 Clones Arrayed on Membranes," *Electrophoresis*, 13:566–573 (1992).

Elkins, et al., "Multianalyte Immunoassay: The Immunological 'Compact Disk' of the Future", *J. Clinical Immunoassay*, 13(4):169–181 (1990).

Primary Examiner—Ardin H. Marschel

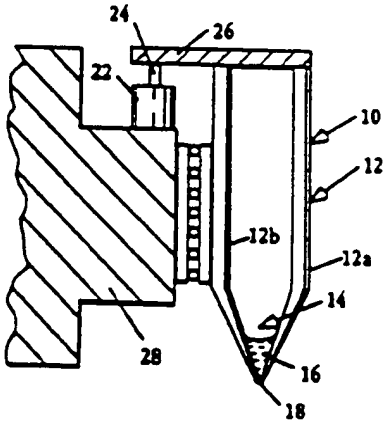
Attorney, Agent, or Firm—Arnold White & Durkee

[57] ABSTRACT

A method and apparatus for forming microarrays of biological samples on a support are disclosed. The method involves dispensing a known volume of a reagent at each selected array position, by tapping a capillary dispenser on the support under conditions effective to draw a defined volume of liquid onto the support. The apparatus is designed to produce a microarray of such regions in an automated fashion.

7 Claims, 6 Drawing Sheets

(2 of 6 Drawing(s) Filed in Color)



U.S. PATENT DOCUMENTS

4,731,325	3/1988	Palva et al.	435/6
4,755,458	7/1988	Rabbani et al.	435/5
4,767,700	8/1988	Wallace	435/6
4,868,104	9/1989	Kurn et al.	435/6
4,868,105	9/1989	Urdea et al.	435/6
4,921,805	5/1990	Gebeyehu et al.	435/270
4,981,783	1/1991	Augenlicht	435/6
5,013,669	5/1991	Peters, Jr. et al.	436/518
5,028,545	7/1991	Smini	436/501
5,064,754	11/1991	Mills	435/6
5,091,652	2/1992	Mathies et al.	250/458.1
5,100,777	3/1992	Chang	435/7.24
5,143,854	9/1992	Pirrung et al.	436/518
5,185,243	2/1993	Ullman et al.	435/6
5,188,963	2/1993	Stapleton	435/288.3
5,200,051	4/1993	Cozzette et al.	204/403
5,200,312	4/1993	Oprandy	435/5
5,202,231	4/1993	Drmanac et al.	435/6
5,204,268	4/1993	Matsumoto	436/44
5,242,974	9/1993	Holmes	525/54.11
5,252,296	10/1993	Zuckerman et al.	422/116
5,252,743	10/1993	Barrett et al.	548/303.7
5,328,824	7/1994	Ward et al.	435/6
5,338,688	8/1994	Dorg et al.	436/180
5,348,855	9/1994	Dattagupta et al.	435/6
5,389,512	2/1995	Soinsky et al.	435/5
5,412,087	5/1995	McGall et al.	536/24.3
5,434,049	7/1995	Okano et al.	435/6
5,445,934	8/1995	Fodor et al.	435/6
5,472,842	12/1995	Stokke et al.	435/6
5,474,796	12/1995	Brennan	427/2.13
5,474,895	12/1995	Ishii et al.	435/6
5,510,270	4/1996	Fodor et al.	436/518
5,512,430	4/1996	Gong	435/5
5,514,543	5/1996	Grossman et al.	435/6
5,514,785	5/1996	Van Ness et al.	536/22.1
5,516,641	5/1996	Ullman et al.	435/6
5,518,883	5/1996	Soini	435/6
5,545,531	8/1996	Rava et al.	435/6
5,556,748	9/1996	Douglas	435/6
5,556,752	9/1996	Lockhart et al.	435/6
5,563,060	10/1996	Hozier	435/240.23
5,578,832	11/1996	Trulson et al.	250/458.1
5,605,662	2/1997	Heller et al.	422/68.1

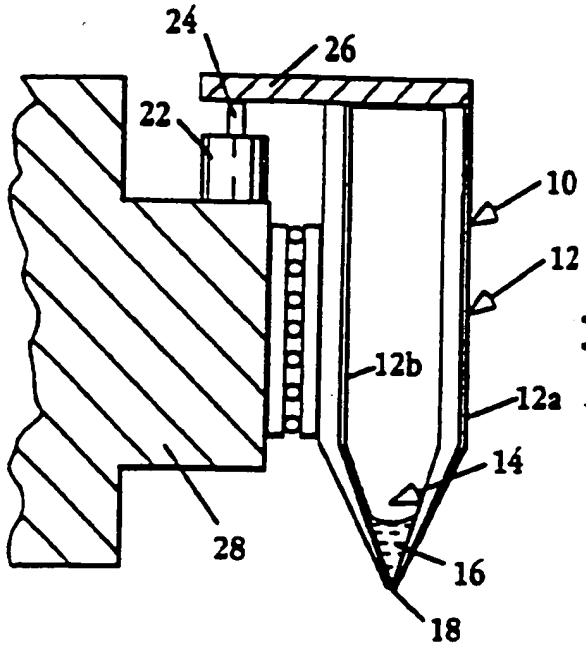


Fig. 1

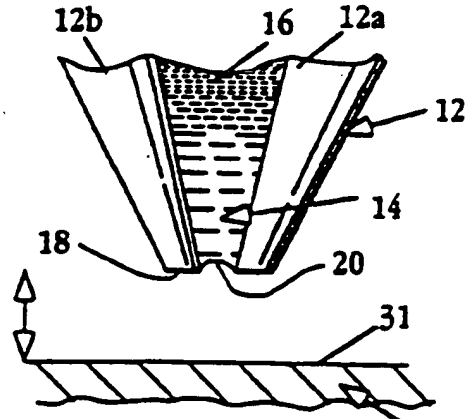


Fig. 2A

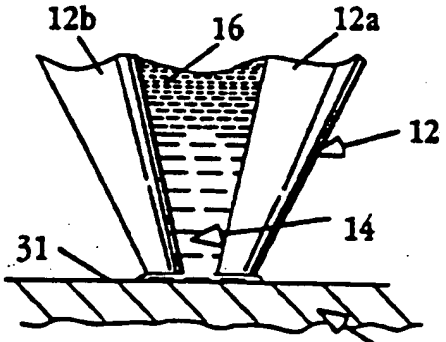


Fig. 2B

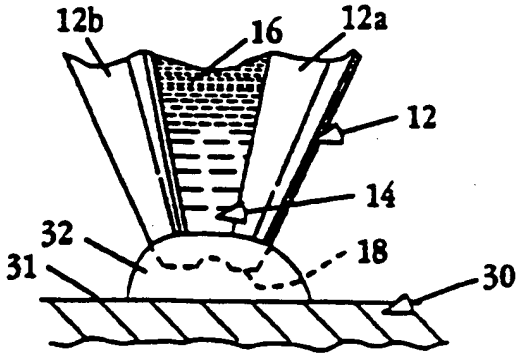


Fig. 2C

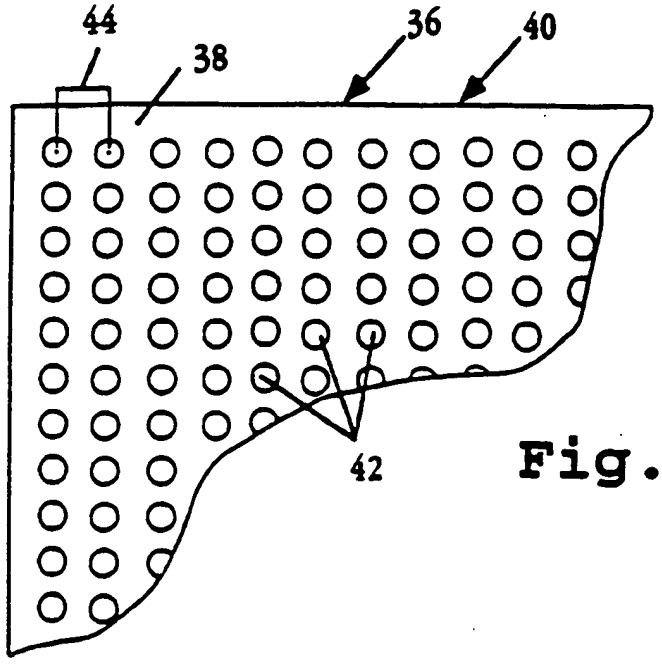

44
38
36
40
42

Fig. 3

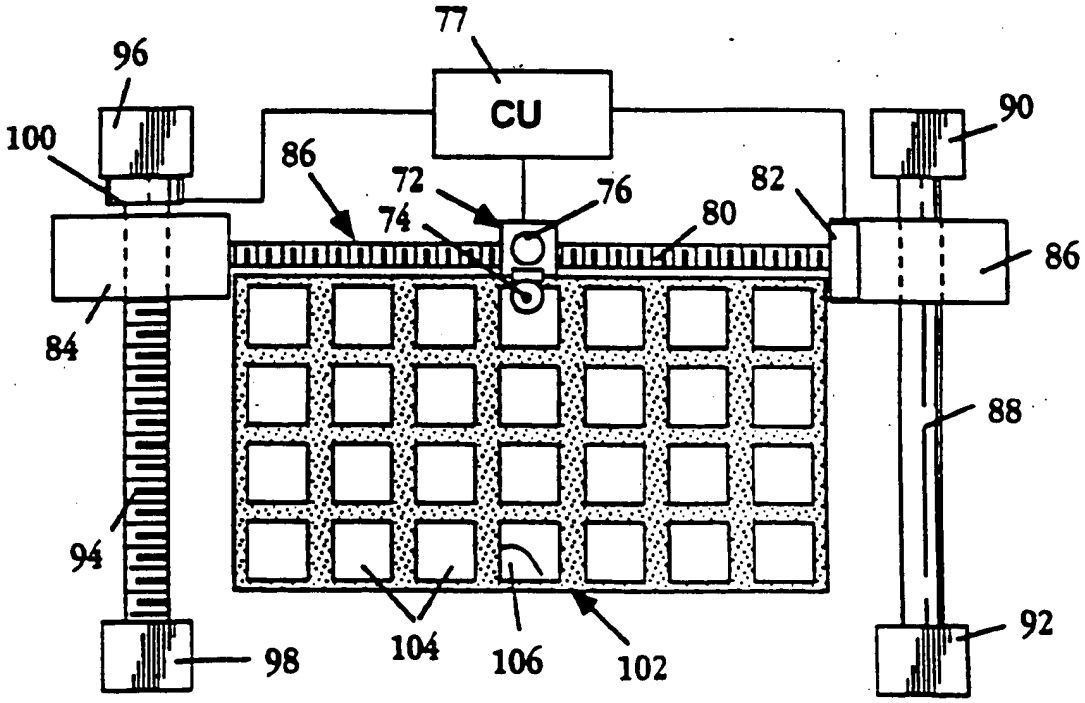

77
CU
96
100
86
72
74
76
80
82
90
86
84
88
94
98
104
106
102
92

Fig. 4

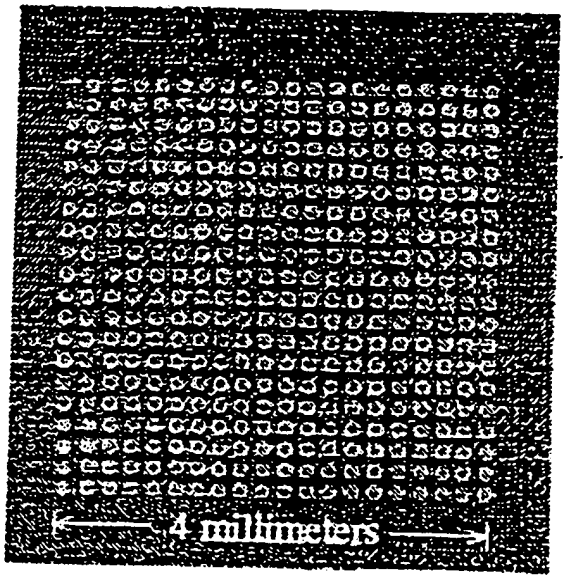


Fig. 5

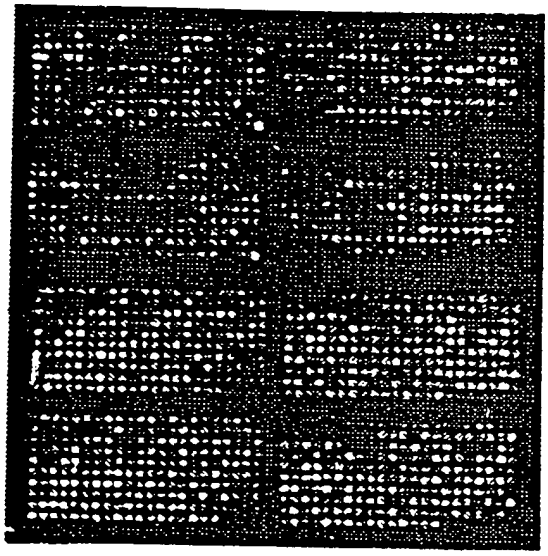


Fig. 6

Fig. 7

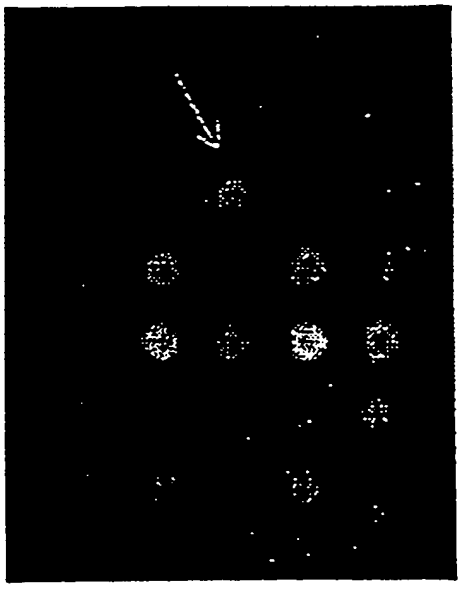


Fig. 8

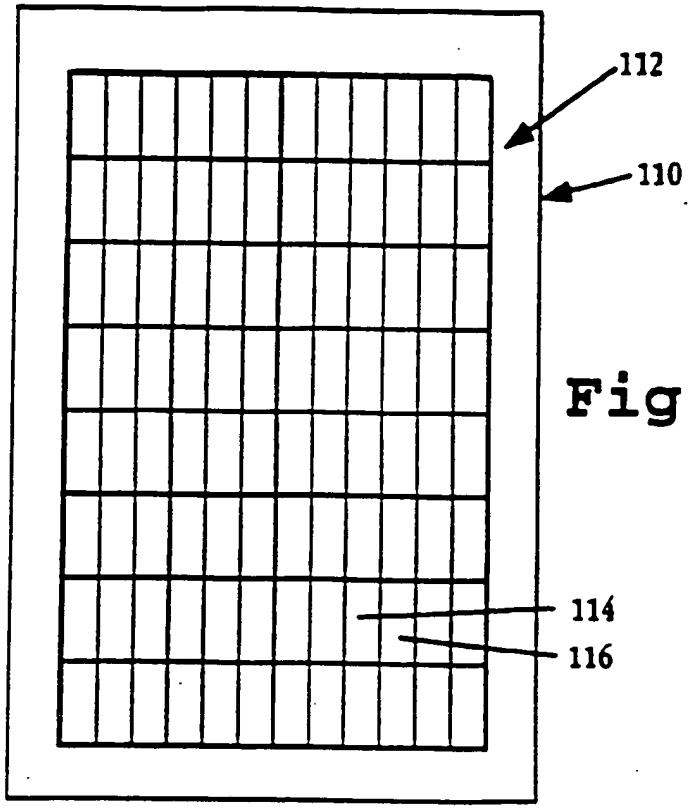


Fig. 9

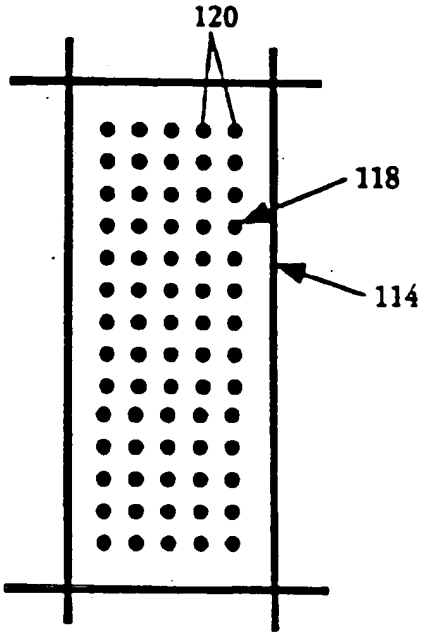


Fig. 10

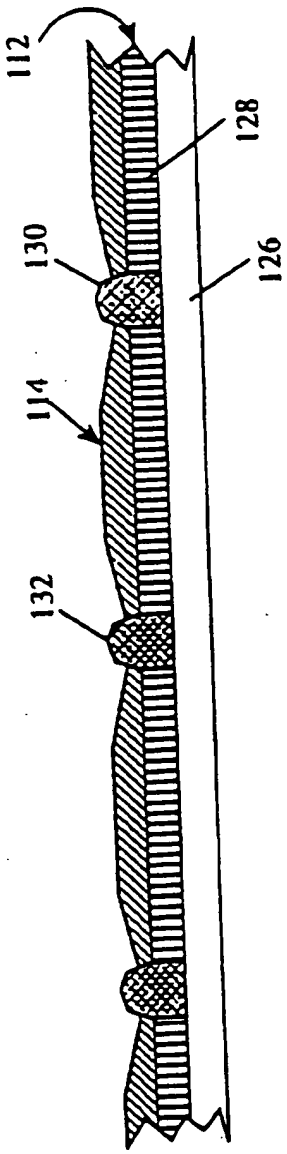


Fig. 11

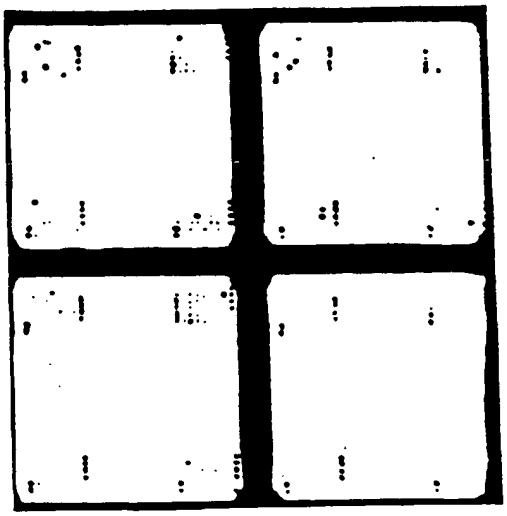


Fig. 12

METHODS FOR FABRICATING MICROARRAYS OF BIOLOGICAL SAMPLES

CROSS-REFERENCE TO RELATED APPLICATION

This application is a continuation-in-part of U.S. patent application Ser. No. 08/261,388, filed Jun. 17, 1994, and now abandoned.

The United States government may have certain rights in the present invention pursuant to Grant No. HG00450 awarded by the National Institutes of Health.

FIELD OF THE INVENTION

This invention relates to a method and apparatus for fabricating microarrays of biological samples for large scale screening assays, such as arrays of DNA samples to be used in DNA hybridization assays for genetic research and diagnostic applications.

REFERENCES

Abouzied, et al., *Journal of AOAC International* 77(2) :495–500 (1994).

Bohlander, et al., *Genomics* 13:1322–1324 (1992).

Drmanac, et al., *Science* 260:1649–1652 (1993).

Fodor, et al., *Science* 251:767–773 (1991).

Khrapko, et al., *DNA Sequence* 1:375–388 (1991).

Kuriyama, et al., *AN ISFET BIOSENSOR, APPLIED BIOSENSORS* (Donald Wise, Ed.), Butterworths, pp. 93–114 (1989).

Lehrach, et al., *HYBRIDIZATION FINGERPRINTING IN GENOME MAPPING AND SEQUENCING, GENOME ANALYSIS*, VOL 1 (Davies and Tilgham, Eds.), Cold Spring Harbor Press, pp. 39–81 (1990).

Maniatis, et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, Cold Spring Harbor Press (1989).

Nelson, et al., *Nature Genetics* 4:11–18 (1993).

Pirrung, et al., U.S. Pat. No. 5,143,854 (1992).

Riles, et al., *Genetics* 134:81–150 (1993).

Schena, M. et al., *Proc. Nat. Acad. Sci. USA* 89:3894–3898 (1992).

Southern, et al., *Genomics* 13:1008–1017 (1992).

BACKGROUND OF THE INVENTION

A variety of methods are currently available for making arrays of biological macromolecules, such as arrays of nucleic acid molecules or proteins. One method for making ordered arrays of DNA on a porous membrane is a "dot blot" approach. In this method, a vacuum manifold transfers a plurality, e.g., 96, aqueous samples of DNA from 3 millimeter diameter wells to a porous membrane. A common variant of this procedure is a "slot-blot" method in which the wells have highly-elongated oval shapes.

The DNA is immobilized on the porous membrane by baking the membrane or exposing it to UV radiation. This is a manual procedure practical for making one array at a time and usually limited to 96 samples per array. "Dot-blot" procedures are therefore inadequate for applications in which many thousand samples must be determined.

A more efficient technique employed for making ordered arrays of genomic fragments uses an array of pins dipped into the wells, e.g., the 96 wells of a microtitre plate, for transferring an array of samples to a substrate, such as a

porous membrane. One array includes pins that are designed to spot a membrane in a staggered fashion, for creating an array of 9216 spots in a 22×22 cm area (Lehrach, et al., 1990). A limitation with this approach is that the volume of DNA spotted in each pixel of each array is highly variable. In addition, the number of arrays that can be made with each dipping is usually quite small.

An alternate method of creating ordered arrays of nucleic acid sequences is described by Pirrung, et al. (1992), and also by Fodor, et al. (1991). The method involves synthesizing different nucleic acid sequences at different discrete regions of a support. This method employs elaborate synthetic schemes, and is generally limited to relatively short nucleic acid sample, e.g., less than 20 bases. A related method has been described by Southern, et al. (1992).

Khrapko, et al. (1991) describes a method of making an oligonucleotide matrix by spotting DNA onto a thin layer of polyacrylamide. The spotting is done manually with a micropipette.

None of the methods or devices described in the prior art are designed for mass fabrication of microarrays characterized by (i) a large number of micro-sized assay regions separated by a distance of 50–200 microns or less, and (ii) a well-defined amount, typically in the picomole range, of analyte associated with each region of the array.

Furthermore, current technology is directed at performing such assays one at a time to a single array of DNA molecules. For example, the most common method for performing DNA hybridizations to arrays spotted onto porous membrane involves sealing the membrane in a plastic bag (Maniatas, et al., 1989) or a rotating glass cylinder (Robbins Scientific) with the labeled hybridization probe inside the sealed chamber. For arrays made on non-porous surfaces, such as a microscope slide, each array is incubated with the labeled hybridization probe sealed under a coverslip. These techniques require a separate sealed chamber for each array which makes the screening and handling of many such arrays inconvenient and time intensive.

Abouzied, et al. (1994) describes a method of printing horizontal lines of antibodies on a nitrocellulose membrane and separating regions of the membrane with vertical stripes of a hydrophobic material. Each vertical stripe is then reacted with a different antigen and the reaction between the immobilized antibody and an antigen is detected using a standard ELISA calorimetric technique. Abouzied's technique makes it possible to screen many one-dimensional arrays simultaneously on a single sheet of nitrocellulose. Abouzied makes the nitrocellulose somewhat hydrophobic using a line drawn with PAP Pen (Research Products International). However, Abouzied does not describe a technology that is capable of completely sealing the pores of the nitrocellulose. The pores of the nitrocellulose are still physically open and so the assay reagents can leak through the hydrophobic barrier during extended high temperature incubations or in the presence of detergents, which makes the Abouzied technique unacceptable for DNA hybridization assays.

Porous membranes with printed patterns of hydrophilic/hydrophobic regions exist for applications such as ordered arrays of bacteria colonies. QA Life Sciences (San Diego Calif.) makes such a membrane with a grid pattern printed on it. However, this membrane has the same disadvantage as the Abouzied technique since reagents can still flow between the gridded arrays making them unusable for separate DNA hybridization assays.

Pall Corporation make a 96-well plate with a porous filter heat sealed to the bottom of the plate. These plates are

capable of containing different reagents in each well without cross-contamination. However, each well is intended to hold only one target element whereas the invention described here makes a microarray of many biomolecules in each subdivided region of the solid support. Furthermore, the 96 well plates are at least 1 cm thick and prevent the use of the device for many calorimetric, fluorescent and radioactive detection formats which require that the membrane lie flat against the detection surface. The invention described here requires no further processing after the assay step since the barriers elements are shallow and do not interfere with the detection step, thereby greatly increasing convenience.

Hyseq Corporation has described a method of making an "array of arrays" on a non-porous solid support for use with their sequencing by hybridization technique. The method described by Hyseq involves modifying the chemistry of the solid support material to form a hydrophobic grid pattern where each subdivided region contains a microarray of biomolecules. Hyseq's flat hydrophobic pattern does not make use of physical blocking as an additional means of preventing cross contamination.

SUMMARY OF THE INVENTION

The invention includes, in one aspect, a method of forming a microarray of analyte-assay regions on a solid support, where each region in the array has a known amount of a selected, analyte-specific reagent. The method involves first loading a solution of a selected analyte-specific reagent in a reagent-dispensing device having an elongate capillary channel (i) formed by spaced-apart, coextensive elongate members, (ii) adapted to hold a quantity of the reagent solution and (iii) having a tip region at which aqueous solution in the channel forms a meniscus. The channel is preferably formed by a pair of spaced-apart tapered elements.

The tip of the dispensing device is tapped against a solid support at a defined position on the support surface with an impulse effective to break the meniscus in the capillary channel, and deposit a selected volume of solution on the surface, preferably a selected volume in the range 0.01 to 100 nl. The two steps are repeated until the desired array is formed.

The method may be practiced in forming a plurality of such arrays, where the solution-depositing step is applied to a selected position on each of a plurality of solid supports at each repeat cycle.

The dispensing device may be loaded with a new solution, by the steps of (i) dipping the capillary channel of the device in a wash solution, (ii) removing wash solution drawn into the capillary channel, and (iii) dipping the capillary channel into the new reagent solution.

Also included in the invention is an automated apparatus for forming a microarray of analyte-assay regions on a plurality of solid supports, where each region in the array has a known amount of a selected, analyte-specific reagent. The apparatus has a holder for holding, at known positions, a plurality of planar supports, and a reagent dispensing device of the type described above.

The apparatus further includes a positioning structure for positioning the dispensing device at a selected array position with respect to a support in said holder, and a dispensing structure for moving the dispensing device into tapping engagement against a support with a selected impulse effective to deposit a selected volume on the support, e.g., a selected volume in the volume range 0.01 to 100 nl.

The positioning and dispensing structures are controlled by a control unit in the apparatus. The unit operates to (i) place the dispensing device at a loading station, (ii) move the capillary channel in the device into a selected reagent at the loading station, to load the dispensing device with the reagent, and (iii) dispense the reagent at a defined array position on each of the supports on said holder. The unit may further operate, at the end of a dispensing cycle, to wash the dispensing device by (i) placing the dispensing device at a washing station, (ii) moving the capillary channel in the device into a wash fluid, to load the dispensing device with the fluid, and (iii) removing the wash fluid prior to loading the dispensing device with a fresh selected reagent.

The dispensing device in the apparatus may be one of a plurality of such devices which are carried on the arm for dispensing different analyte assay reagents at selected spaced array positions.

In another aspect, the invention includes a substrate with a surface having a microarray of at least 10^3 distinct polynucleotide or polypeptide biopolymers in a surface area of less than about 1 cm^2. Each distinct biopolymer (i) is disposed at a separate, defined position in said array, (ii) has a length of at least 50 subunits, and (iii) is present in a defined amount between about 0.1 femtomoles and 100 nanomoles.

In one embodiment, the surface is glass slide surface coated with a polycationic polymer, such as polylysine, and the biopolymers are polynucleotides. In another embodiment, the substrate has a water-impermeable backing, a water-permeable film formed on the backing, and a grid formed on the film. The grid is composed of intersecting water-impervious grid elements extending from said backing to positions raised above the surface of said film, and partitions the film into a plurality of water-impervious cells. A biopolymer array is formed within each well.

More generally, there is provided a substrate for use in detecting binding of labeled polynucleotides to one or more of a plurality different-sequence, immobilized polynucleotides. The substrate includes, in one aspect, a glass support, a coating of a polycationic polymer, such as polylysine, on said surface of the support, and an array of distinct polynucleotides electrostatically bound non-covalently to said coating, where each distinct biopolymer is disposed at a separate, defined position in a surface array of polynucleotides.

In another aspect, the substrate includes a water-impermeable backing, a water-permeable film formed on the backing, and a grid formed on the film, where the grid is composed of intersecting water-impervious grid elements extending from the backing to positions raised above the surface of the film, forming a plurality of cells. A biopolymer array is formed within each cell.

Also forming part of the invention is a method of detecting differential expression of each of a plurality of genes in a first cell type, with respect to expression of the same genes in a second cell type. In practicing the method, there is first produced fluorescent-labeled cDNAs from mRNAs isolated from the two cells types, where the cDNAs from the first and second cell types are labeled with first and second different fluorescent reporters.

A mixture of the labeled cDNAs from the two cell types is added to an array of polynucleotides representing a plurality of known genes derived from the two cell types, under conditions that result in hybridization of the cDNAs to complementary-sequence polynucleotides in the array. The array is then examined by fluorescence under fluorescence excitation conditions in which (i) polynucleotides in the array that are hybridized predominantly to cDNAs derived

from one of the first or second cell types give a distinct first or second fluorescence emission color, respectively, and (ii) polynucleotides in the array that are hybridized to substantially equal numbers of cDNAs derived from the first and second cell types give a distinct combined fluorescence emission color, respectively. The relative expression of known genes in the two cell types can then be determined by the observed fluorescence emission color of each spot.

These and other objects and features of the invention will become more fully apparent when the following detailed description of the invention is read in conjunction with the accompanying figures.

The file of this patent contains at least one drawing executed in color. Copies of this patent with color drawing (s) will be provided by the Patent and Trademark Office upon request and payment of the necessary fee.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a side view of a reagent-dispensing device having a open-capillary dispensing head constructed for use in one embodiment of the invention;

FIGS. 2A–2C illustrate steps in the delivery of a fixed-volume bead on a hydrophobic surface employing the dispensing head from FIG. 1, in accordance with one embodiment of the method of the invention;

FIG. 3 shows a portion of a two-dimensional array of analyte-assay regions constructed according to the method of the invention;

FIG. 4 is a planar view showing components of an automated apparatus for forming arrays in accordance with the invention.

FIG. 5 shows a fluorescent image of an actual 20×20 array of 400 fluorescently-labeled DNA samples immobilized on a poly-l-lysine coated slide, where the total area covered by the 400 element array is 16 square millimeters;

FIG. 6 is a fluorescent image of a 1.8 cm×1.8 cm microarray containing lambda clones with yeast inserts, the fluorescent signal arising from the hybridization to the array with approximately half the yeast genome labeled with a green fluorophore and the other half with a red fluorophore;

FIG. 7 shows the translation of the hybridization image of FIG. 6 into a karyotype of the yeast genome, where the elements of FIG. 6 microarray contain yeast DNA sequences that have been previously physically mapped in the yeast genome;

FIG. 8 shows a fluorescent image of a 0.5 cm×0.5 cm microarray of 24 cDNA clones, where the microarray was hybridized simultaneously with total cDNA from wild type Arabidopsis plant labeled with a green fluorophore and total cDNA from a transgenic Arabidopsis plant labeled with a red fluorophore, and the arrow points to the cDNA clone representing the gene introduced into the transgenic Arabidopsis plant;

FIG. 9 shows a plan view of substrate having an array of cells formed by barrier elements in the form of a grid;

FIG. 10 shows an enlarged plan view of one of the cells in the substrate in FIG. 9, showing an array of polynucleotide regions in the cell;

FIG. 11 is an enlarged sectional view of the substrate in FIG. 9, taken along a section line in that figure; and

FIG. 12 is a scanned image of a 3 cm×3 cm nitrocellulose solid support containing four identical arrays of M13 clones in each of four quadrants, where each quadrant was hybridized simultaneously to a different oligonucleotide using an open face hybridization method.

DETAILED DESCRIPTION OF THE INVENTION

I. Definitions

Unless indicated otherwise, the terms defined below have the following meanings:

"Ligand" refers to one member of a ligand/anti-ligand binding pair. The ligand may be, for example, one of the nucleic acid strands in a complementary, hybridized nucleic acid duplex binding pair; an effector molecule in an effector/receptor binding pair; or an antigen in an antigen/antibody or antigen/antibody fragment binding pair.

"Anti-ligand" refers to the opposite member of a ligand/anti-ligand binding pair. The anti-ligand may be the other of the nucleic acid strands in a complementary, hybridized nucleic acid duplex binding pair; the receptor molecule in an effector/receptor binding pair; or an antibody or antibody fragment molecule in antigen/antibody or antigen/antibody fragment binding pair, respectively.

"Analyte" or "analyte molecule" refers to a molecule, typically a macromolecule, such as a polynucleotide or polypeptide, whose presence, amount, and/or identity are to be determined. The analyte is one member of a ligand/anti-ligand pair.

"Analyte-specific assay reagent" refers to a molecule effective to bind specifically to an analyte molecule. The reagent is the opposite member of a ligand/anti-ligand binding pair.

An "array of regions on a solid support" is a linear or two-dimensional array of preferably discrete regions, each having a finite area, formed on the surface of a solid support.

A "microarray" is an array of regions having a density of discrete regions of at least about 100/cm^2, and preferably at least about 1000/cm^2. The regions in a microarray have typical dimensions, e.g., diameters, in the range of between about 10–250 μm, and are separated from other regions in the array by about the same distance.

A support surface is "hydrophobic" if a aqueous-medium droplet applied to the surface does not spread out substantially beyond the area size of the applied droplet. That is, the surface acts to prevent spreading of the droplet applied to the surface by hydrophobic interaction with the droplet.

A "meniscus" means a concave or convex surface that forms on the bottom of a liquid in a channel as a result of the surface tension of the liquid.

"Distinct biopolymers", as applied to the biopolymers forming a microarray, means an array member which is distinct from other array members on the basis of a different biopolymer sequence, and/or different concentrations of the same or distinct biopolymers, and/or different mixtures of distinct or different-concentration biopolymers. Thus an array of "distinct polynucleotides" means an array containing, as its members, (i) distinct polynucleotides, which may have a defined amount in each member, (ii) different, graded concentrations of given-sequence polynucleotides, and/or (iii) different-composition mixtures of two or more distinct polynucleotides.

"Cell type" means a cell from a given source, e.g., a tissue, or organ, or a cell in a given state of differentiation, or a cell associated with a given pathology or genetic makeup.

II. Method of Microarray Formation

This section describes a method of forming a microarray of analyte-assay regions on a solid support or substrate, where each region in the array has a known amount of a selected, analyte-specific reagent.

FIG. 1 illustrates, in a partially schematic view, a reagent-dispensing device 10 useful in practicing the method. The device generally includes a reagent dispenser 12 having an elongate open capillary channel 14 adapted to hold a quantity of the reagent solution, such as indicated at 16, as will be described below. The capillary channel is formed by a pair of spaced-apart, coextensive, elongate members 12a, 12b which are tapered toward one another and converge at a tip or tip region 18 at the lower end of the channel. More generally, the open channel is formed by at least two elongate, spaced-apart members adapted to hold a quantity of reagent solutions and having a tip region at which aqueous solution in the channel forms a meniscus, such as the concave meniscus illustrated at 20 in FIG. 2A. The advantages of the open channel construction of the dispenser are discussed below.

With continued reference to FIG. 1, the dispenser device also includes structure for moving the dispenser rapidly toward and away from a support surface, for effecting deposition of a known amount of solution in the dispenser on a support, as will be described below with reference to FIGS. 2A–2C. In the embodiment shown, this structure includes a solenoid 22 which is activatable to draw a solenoid piston 24 rapidly downwardly, then release the piston, e.g., under spring bias, to a normal, raised position, as shown. The dispenser is carried on the piston by a connecting member 26, as shown. The just-described moving structure is also referred to herein as dispensing means for moving the dispenser into engagement with a solid support, for dispensing a known volume of fluid on the support.

The dispensing device just described is carried on an arm 28 that may be moved either linearly or in an x-y plane to position the dispenser at a selected deposition position, as will be described.

FIGS. 2A–2C illustrate the method of depositing a known amount of reagent solution in the just-described dispenser on the surface of a solid support, such as the support indicated at 30. The support is a polymer, glass, or other solid-material support having a surface indicated at 31.

In one general embodiment, the surface is a relatively hydrophilic, i.e., wettable surface, such as a surface having native, bound or covalently attached charged groups. One such surface described below is a glass surface having an absorbed layer of a polycationic polymer, such as poly-l-lysine.

In another embodiment, the surface has or is formed to have a relatively hydrophobic character, i.e., one that causes aqueous medium deposited on the surface to bead. A variety of known hydrophobic polymers, such as polystyrene, polypropylene, or polyethylene have desired hydrophobic properties, as do glass and a variety of lubricant or other hydrophobic films that may be applied to the support surface.

Initially, the dispenser is loaded with a selected analyte-specific reagent solution, such as by dipping the dispenser tip, after washing, into a solution of the reagent, and allowing filling by capillary flow into the dispenser channel. The dispenser is now moved to a selected position with respect to a support surface, placing the dispenser tip directly above the support-surface position at which the reagent is to be deposited. This movement takes place with the dispenser tip in its raised position, as seen in FIG. 2A, where the tip is typically at least several 1–5 mm above the surface of the substrate.

With the dispenser so positioned, solenoid 22 is now activated to cause the dispenser tip to move rapidly toward and away from the substrate surface, making momentary contact with the surface, in effect, tapping the tip of the dispenser against the support surface. The tapping movement of the tip against the surface acts to break the liquid meniscus in the tip channel, bringing the liquid in the tip into contact with the support surface. This, in turn, produces a flowing of the liquid into the capillary space between the tip and the surface, acting to draw liquid out of the dispenser channel, as seen in FIG. 2B.

FIG. 2C shows flow of fluid from the tip onto the support surface, which in this case is a hydrophobic surface. The figure illustrates that liquid continues to flow from the dispenser onto the support surface until it forms a liquid bead 32. At a given bead size, i.e., volume, the tendency of liquid to flow onto the surface will be balanced by the hydrophobic surface interaction of the bead with the support surface, which acts to limit the total bead area on the surface, and by the surface tension of the droplet, which tends toward a given bead curvature. At this point, a given bead volume will have formed, and continued contact of the dispenser tip with the bead, as the dispenser tip is being withdrawn, will have little or no effect on bead volume.

For liquid-dispensing on a more hydrophilic surface, the liquid will have less of a tendency to bead, and the dispensed volume will be more sensitive to the total dwell time of the dispenser tip in the immediate vicinity of the support surface, e.g., the positions illustrated in FIGS. 2B and 2C.

The desired deposition volume, i.e., bead volume, formed by this method is preferably in the range 2 pl (picoliters) to 2 nl (nanoliters), although volumes as high as 100 nl or more may be dispensed. It will be appreciated that the selected dispensed volume will depend on (i) the "footprint" of the dispenser tip, i.e., the size of the area spanned by the tip, (ii) the hydrophobicity of the support surface, and (iii) the time of contact with and rate of withdrawal of the tip from the support surface. In addition, bead size may be reduced by increasing the viscosity of the medium, effectively reducing the flow time of liquid from the dispenser onto the support surface. The drop size may be further constrained by depositing the drop in a hydrophilic region surrounded by a hydrophobic grid pattern on the support surface.

In a typical embodiment, the dispenser tip is tapped rapidly against the support surface, with a total resilience time in contact with the support of less than about 1 msec, and a rate of upward travel from the surface of about 10 cm/sec.

Assuming that the bead that forms on contact with the surface is a hemispherical head, with a diameter approximately equal to the width of the dispenser tip, as shown in FIG. 2C, the volume of the bead formed in relation to dispenser tip width (d) is given in Table 1 below. As seen, the volume of the bead ranges between 2 pl to 2 nl as the width size is increased from about 20 to 200 μm.

TABLE 1

d	Volume (nl)
20 μm	2×10^{-3}
50 μm	3.1×10^{-2}
100 μm	2.5×10^{-1}
200 μm	2

At a given tip size, bead volume can be reduced in a controlled fashion by increasing surface hydrophobicity, reducing time of contact of the tip with the surface, increasing rate of movement of the tip away from the surface,

and/or increasing the viscosity of the medium. Once these parameters are fixed, a selected deposition volume in the desired pl to nl range can be achieved in a repeatable fashion.

After depositing a bead at one selected location on a support, the tip is typically moved to a corresponding position on a second support, a droplet is deposited at that position, and this process is repeated until a liquid droplet of the reagent has been deposited at a selected position on each of a plurality of supports.

The tip is then washed to remove the reagent liquid, filled with another reagent liquid and this reagent is now deposited at each another array position on each of the supports. In one embodiment, the tip is washed and refilled by the steps of (i) dipping the capillary channel of the device in a wash solution, (ii) removing wash solution drawn into the capillary channel, and (iii) dipping the capillary channel into the new reagent solution.

From the foregoing, it will be appreciated that the tweezers-like, open-capillary dispenser tip provides the advantages that (i) the open channel of the tip facilitates rapid, efficient washing and drying before reloading the tip with a new reagent, (ii) passive capillary action can load the sample directly from a standard microwell plate while retaining sufficient sample in the open capillary reservoir for the printing of numerous arrays, (iii) open capillaries are less prone to clogging than closed capillaries, and (iv) open capillaries do not require a perfectly faced bottom surface for fluid delivery.

A portion of a microarray 36 formed on the surface 38 of a solid support 40 in accordance with the method just described is shown in FIG. 3. The array is formed of a plurality of analyte-specific reagent regions, such as regions 42, where each region may include a different analyte-specific reagent. As indicated above, the diameter of each region is preferably between about 20–200 μm. The spacing between each region and its closest (non-diagonal) neighbor, measured from center-to-center (indicated at 44), is preferably in the range of about 20–400 μm. Thus, for example, an array having a center-to-center spacing of about 250 μm contains about 40 regions/cm or 1,600 regions/cm^2. After formation of the array, the support is treated to evaporate the liquid of the droplet forming each region, to leave a desired array of dried, relatively flat regions. This drying may be done by heating or under vacuum.

In some cases, it is desired to first rehydrate the droplets containing the analyte reagents to allow for more time for adsorption to the solid support. It is also possible to spot out the analyte reagents in a humid environment so that droplets do not dry until the arraying operation is complete.

III. Automated Apparatus for Forming Arrays

In another aspect, the invention includes an automated apparatus for forming an array of analyte-assay regions on a solid support, where each region in the array has a known amount of a selected, analyte-specific reagent.

The apparatus is shown in planar, and partially schematic view in FIG. 4. A dispenser device 72 in the apparatus has the basic construction described above with respect to FIG. 1, and includes a dispenser 74 having an open-capillary channel terminating at a tip, substantially as shown in FIGS. 1 and 2A–2C.

The dispenser is mounted in the device for movement toward and away from a dispensing position at which the tip of the dispenser taps a support surface, to dispense a selected volume of reagent solution, as described above. This movement is effected by a solenoid 76 as described above. Solenoid 76 is under the control of a control unit 77 whose operation will be described below. The solenoid is also referred to herein as dispensing means for moving the device into tapping engagement with a support, when the device is positioned at a defined array position with respect to that support.

The dispenser device is carried on an arm 74 which is threadedly mounted on a worm screw 80 driven (rotated) in a desired direction by a stepper motor 82 also under the control of unit 77. At its left end in the figure screw 80 is carried in a sleeve 84 for rotation about the screw axis. At its other end, the screw is mounted to the drive shaft of the stepper motor, which in turn is carried on a sleeve 86. The dispenser device, worm screw, the two sleeves mounting the worm screw, and the stepper motor used in moving the device in the "x" (horizontal) direction in the figure form what is referred to here collectively as a displacement assembly 86.

The displacement assembly is constructed to produce precise, micro-range movement in the direction of the screw, i.e., along an x axis in the figure. In one mode, the assembly functions to move the dispenser in x-axis increments having a selected distance in the range 5–25 μm. In another mode, the dispenser unit may be moved in precise x-axis increments of several microns or more, for positioning the dispenser at associated positions on adjacent supports, as will be described below.

The displacement assembly, in turn, is mounted for movement in the "y" (vertical) axis of the figure, for positioning the dispenser at a selected y axis position. The structure mounting the assembly includes a fixed rod 88 mounted rigidly between a pair of frame bars 90, 92, and a worm screw 94 mounted for rotation between a pair of frame bars 96, 98. The worm screw is driven (rotated) by a stepper motor 100 which operates under the control of unit 77. The motor is mounted on bar 96, as shown.

The structure just described, including worm screw 94 and motor 100, is constructed to produce precise, micro-range movement in the direction of the screw, i.e., along a y axis in the figure. As above, the structure functions in one mode to move the dispenser in y-axis increments having a selected distance in the range 5–250 μm, and in a second mode, to move the dispenser in precise y-axis increments of several microns (μm) or more, for positioning the dispenser at associated positions on adjacent supports.

The displacement assembly and structure for moving this assembly in the y axis are referred to herein collectively as positioning means for positioning the dispensing device at a selected array position with respect to a support.

A holder 102 in the apparatus functions to hold a plurality of supports, such as supports 104 on which the microarrays of reagent regions are to be formed by the apparatus. The holder provides a number of recessed slots, such as slot 106, which receive the supports, and position them at precise selected positions with respect to the frame bars on which the dispenser moving means is mounted.

As noted above, the control unit in the device functions to actuate the two stepper motors and dispenser solenoid in a sequence designed for automated operation of the apparatus in forming a selected microarray of reagent regions on each of a plurality of supports.

The control unit is constructed, according to conventional microprocessor control principles, to provide appropriate signals to each of the solenoid and each of the stepper motors, in a given timed sequence and for appropriate signalling time. The construction of the unit, and the settings

that are selected by the user to achieve a desired array pattern, will be understood from the following description of a typical apparatus operation.

Initially, one or more supports are placed in one or more slots in the holder. The dispenser is then moved to a position directly above a well (not shown) containing a solution of the first reagent to be dispensed on the support(s). The dispenser solenoid is actuated now to lower the dispenser tip into this well, causing the capillary channel in the dispenser to fill. Motors 82, 100 are now actuated to position the dispenser at a selected array position at the first of the supports. Solenoid actuation of the dispenser is then effective to dispense a selected-volume droplet of that reagent at this location. As noted above, this operation is effective to dispense a selected volume preferably between 2 pl and 2 nl of the reagent solution.

The dispenser is now moved to the corresponding position at an adjacent support and a similar volume of the solution is dispensed at this position. The process is repeated until the reagent has been dispensed at this preselected corresponding position on each of the supports.

Where it is desired to dispense a single reagent at more than two array positions on a support, the dispenser may be moved to different array positions at each support, before moving the dispenser to a new support, or solution can be dispensed at individual positions on each support, at one selected position, then the cycle repeated for each new array position.

To dispense the next reagent, the dispenser is positioned over a wash solution (not shown), and the dispenser tip is dipped in and out of this solution until the reagent solution has been substantially washed from the tip. Solution can be removed from the tip, after each dipping, by vacuum, compressed air spray, sponge, or the like.

The dispenser tip is now dipped in a second reagent well, and the filled tip is moved to a second selected array position in the first support. The process of dispensing reagent at each of the corresponding second-array positions is then carried out as above. This process is repeated until an entire microarray of reagent solutions on each of the supports has been formed.

IV. Microarray Substrate

This section describes embodiments of a substrate having a microarray of biological polymers carried on the substrate surface. Subsection A describes a multi-cell substrate, each cell of which contains a microarray, and preferably an identical microarray, of distinct biopolymers, such as distinct polynucleotides, formed on a porous surface. Subsection B describes a microarray of distinct polynucleotides bound on a glass slide coated with a polycationic polymer.

A. Multi-Cell Substrate

FIG. 9 illustrates, in plan view, a substrate 110 constructed according to the invention. The substrate has an 8×12 rectangular array 112 of cells, such as cells 114, 116, formed on the substrate surface. With reference to FIG. 10, each cell, such as cell 114, in turn supports a microarray 118 of distinct biopolymers, such as polypeptides or polynucleotides at known, addressable regions of the microarray. Two such regions forming the microarray are indicated at 120, and correspond to regions, such as regions 42, forming the microarray of distinct biopolymers shown in FIG. 3.

The 96-cell array shown in FIG. 9 typically has array dimensions between about 12 and 244 mm in width and 8 and 400 mm in length, with the cells in the array having width and length dimension of 1/12 and 1/8 the array width and length dimensions, respectively, i.e., between about 1 and 20 in width and 1 and 50 mm in length.

The construction of substrate is shown cross-sectionally in FIG. 11, which is an enlarged sectional view taken along view line 124 in FIG. 9. The substrate includes a water-impermeable backing 126, such as a glass slide or rigid polymer sheet. Formed on the surface of the backing is a water-permeable film 128. The film is formed of a porous membrane material, such as nitrocellulose membrane, or a porous web material, such as a nylon, polypropylene, or PVDF porous polymer material. The thickness of the film is preferably between about 10 and 1000 μm. The film may be applied to the backing by spraying or coating uncured material on the backing, or by applying a preformed membrane to the backing. The backing and film may be obtained as a preformed unit from commercial source, e.g., a plastic-backed nitrocellulose film available from Schleicher and Schuell Corporation.

With continued reference to FIG. 11, the film-covered surface in the substrate is partitioned into a desired array of cells by water-impermeable grid lines, such as lines 130, 132, which have infiltrated the film down to the level of the backing, and extend above the surface of the film as shown, typically a distance of 100 to 2000 μm above the film surface.

The grid lines are formed on the substrate by laying down an uncured or otherwise flowable resin or elastomer solution in an array grid, allowing the material to infiltrate the porous film down to the backing, then curing or otherwise hardening the grid lines to form the cell-array substrate.

One preferred material for the grid is a flowable silicone available from Loctite Corporation. The barrier material can be extruded through a narrow syringe (e.g., 22 gauge) using air pressure or mechanical pressure. The syringe is moved relative to the solid support to print the barrier elements as a grid pattern. The extruded bead of silicone wicks into the pores of the solid support and cures to form a shallow waterproof barrier separating the regions of the solid support.

In alternative embodiments, the barrier element can be a wax-based material or a thermoset material such as epoxy. The barrier material can also be a UV-curing polymer which is exposed to UV light after being printed onto the solid support. The barrier material may also be applied to the solid support using printing techniques such as silk-screen printing. The barrier material may also be a heat-seal stamping of the porous solid support which seals its pores and forms a water-impervious barrier element. The barrier material may also be a shallow grid which is laminated or otherwise adhered to the solid support.

In addition to plastic-backed nitrocellulose, the solid support can be virtually any porous membrane with or without a non-porous backing. Such membranes are readily available from numerous vendors and are made from nylon, PVDF, polysulfone and the like. In an alternative embodiment, the barrier element may also be used to adhere the porous membrane to a non-porous backing in addition to functioning as a barrier to prevent cross contamination of the assay reagents.

In an alternative embodiment, the solid support can be of a non-porous material. The barrier can be printed either before or after the microarray of biomolecules is printed on the solid support.

As can be appreciated, the cells formed by the grid lines and the underlying backing are water-impermeable, having side barriers projecting above the porous film in the cells. Thus, defined-volume samples can be placed in each well without risk of cross-contamination with sample material in adjacent cells. In FIG. 11, defined volumes samples, such as sample 134, are shown in the cells.

As noted above, each well contains a microarray of distinct biopolymers. In one general embodiment, the microarrays in the well are identical arrays of distinct biopolymers, e.g., different sequence polynucleotides. Such arrays can be formed in accordance with the methods described in Section II, by depositing a first selected polynucleotide at the same selected microarray position in each of the cells, then depositing a second polynucleotide at a different microarray position in each well, and so on until a complete, identical microarray is formed in each cell.

In a preferred embodiment, each microarray contains about 10^3 distinct polynucleotide or polypeptide biopolymers per surface area of less than about 1 cm^2. Also in a preferred embodiment, the biopolymers in each microarray region are present in a defined amount between about 0.1 femtomoles and 100 nanomoles. The ability to form high-density arrays of biopolymers, where each region is formed of a well-defined amount of deposited material, can be achieved in accordance with the microarray-forming method described in Section II.

Also in a preferred embodiment, the biopolymers are polynucleotides having lengths of at least about 50 bp, i.e., substantially longer than oligonucleotides which can be formed in high-density arrays by schemes involving parallel, step-wise polymer synthesis on the array surface.

In the case of a polynucleotide array, in an assay procedure, a small volume of the labeled DNA probe mixture in a standard hybridization solution is loaded onto each cell. The solution will spread to cover the entire microarray and stop at the barrier elements. The solid support is then incubated in a humid chamber at the appropriate temperature as required by the assay.

Each assay may be conducted in an "open-face" format where no further sealing step is required, since the hybridization solution will be kept properly hydrated by the water vapor in the humid chamber. At the conclusion of the incubation step, the entire solid support containing the numerous microarrays is rinsed quickly enough to dilute the assay reagents so that no significant cross contamination occurs. The entire solid support is then reacted with detection reagents if needed and analyzed using standard calorimetric, radioactive or fluorescent detection means. All processing and detection steps are performed simultaneously to all of the microarrays on the solid support ensuring uniform assay conditions for all of the microarrays on the solid support.

B. Glass-Slide Polynucleotide Array

FIG. 5 shows a substrate 136 formed according to another aspect of the invention, and intended for use in detecting binding of labeled polynucleotides to one or more of a plurality distinct polynucleotides. The substrate includes a glass substrate 138 having formed on its surface, a coating of a polycationic polymer, preferably a cationic polypeptide, such as polylysine or polyarginine. Formed on the polycationic coating is a microarray 140 of distinct polynucleotides, each localized at known selected array regions, such as regions 142.

The slide is coated by placing a uniform-thickness film of a polycationic polymer, e.g., poly-l-lysine, on the surface of a slide and drying the film to form a dried coating. The amount of polycationic polymer added is sufficient to form at least a monolayer of polymers on the glass surface. The polymer film is bound to surface via electrostatic binding between negative silyl-OH groups on the surface and charged amine groups in the polymers. Poly-l-lysine coated glass slides may be obtained commercially, e.g., from Sigma Chemical Co. (St. Louis, Mo.).

To form the microarray, defined volumes of distinct polynucleotides are deposited on the polymer-coated slide, as described in Section II. According to an important feature of the substrate, the deposited polynucleotides remain bound to the coated slide surface non-covalently when an aqueous DNA sample is applied to the substrate under conditions which allow hybridization of reporter-labeled polynucleotides in the sample to complementary-sequence (single-stranded) polynucleotides in the substrate array. The method is illustrated in Examples 1 and 2.

To illustrate this feature, a substrate of the type just described, but having an array of same-sequence polynucleotides, was mixed with fluorescent-labeled complementary DNA under hybridization conditions. After washing to remove non-hybridized material, the substrate was examined by low-power fluorescence microscopy. The array can be visualized by the relatively uniform labeling pattern of the array regions.

In a preferred embodiment, each microarray contains at least 10^3 distinct polynucleotide or polypeptide biopolymers per surface area of less than about 1 cm^2. In the embodiment shown in FIG. 5, the microarray contains 400 regions in an area of about 16 mm^2, or 2.5×10^3 regions/cm^2. Also in a preferred embodiment, the polynucleotides in each microarray region are present in a defined amount between about 0.1 femtomoles and 100 nanomoles in the case of polynucleotides. As above, the ability to form high-density arrays of this type, where each region is formed of a well-defined amount of deposited material, can be achieved in accordance with the microarray-forming method described in Section II.

Also in a preferred embodiment, the polynucleotides have lengths of at least about 50 bp, i.e., substantially longer than oligonucleotides which can be formed in high-density arrays by various in situ synthesis schemes.

V. Utility

Microarrays of immobilized nucleic acid sequences prepared in accordance with the invention can be used for large scale hybridization assays in numerous genetic applications, including genetic and physical mapping of genomes, monitoring of gene expression, DNA sequencing, genetic diagnosis, genotyping of organisms, and distribution of DNA reagents to researchers.

For gene mapping, a gene or a cloned DNA fragment is hybridized to an ordered array of DNA fragments, and the identity of the DNA elements applied to the array is unambiguously established by the pixel or pattern of pixels of the array that are detected. One application of such arrays for creating a genetic map is described by Nelson, et al. (1993). In constructing physical maps of the genome, arrays of immobilized cloned DNA fragments are hybridized with other cloned DNA fragments to establish whether the cloned fragments in the probe mixture overlap and are therefore contiguous to the immobilized clones on the array. For example, Lehrach, et al., describe such a process.

The arrays of immobilized DNA fragments may also be used for genetic diagnostics. To illustrate, an array containing multiple forms of a mutated gene or genes can be probed with a labeled mixture of a patient's DNA which will preferentially interact with only one of the immobilized versions of the gene.

The detection of this interaction can lead to a medical diagnosis. Arrays of immobilized DNA fragments can also be used in DNA probe diagnostics. For example, the identity of a pathogenic microorganism can be established unambiguously by hybridizing a sample of the unknown pathogen's DNA to an array containing many types of known pathogenic DNA. A similar technique can also be used for

unambiguous genotyping of any organism. Other molecules of genetic interest, such as cDNAs and RNAs can be immobilized on the array or alternately used as the labeled probe mixture that is applied to the array.

In one application, an array of cDNA clones representing genes is hybridized with total cDNA from an organism to monitor gene expression for research or diagnostic purposes. Labeling total cDNA from a normal cell with one color fluorophore and total cDNA from a diseased cell with another color fluorophore and simultaneously hybridizing the two cDNA samples to the same array of cDNA clones allows for differential gene expression to be measured as the ratio of the two fluorophore intensities. This two-color experiment can be used to monitor gene expression in different tissue types, disease states, response to drugs, or response to environmental factors. An example of this approach is illustrated in Example 2, described with respect to FIG. 8.

By way of example and without implying a limitation of scope, such a procedure could be used to simultaneously screen many patients against all known mutations in a disease gene. This invention could be used in the form of, for example, 96 identical 0.9 cmx2.2 cm microarrays fabricated on a single 12 cmx18 cm sheet of plastic-backed nitrocellulose where each microarray could contain, for example, 100 DNA fragments representing all known mutations of a given gene. The region of interest from each of the DNA samples from 96 patients could be amplified, labeled, and hybridized to the 96 individual arrays with each assay performed in 100 microliters of hybridization solution. The approximately 1 thick silicone rubber barrier elements between individual arrays prevent cross-contamination of the patient samples by sealing the pores of the nitrocellulose and by acting as a physical barrier between each microarray. The solid support containing all 96 microarrays assayed with the 96 patient samples is incubated, rinsed, detected and analyzed as a single sheet of material using standard radioactive, fluorescent, or colorimetric detection means (Maniatas, et al., 1989). Previously, such a procedure would involve the handling, processing and tracking of 96 separate membranes in 96 separate sealed chambers. By processing all 96 arrays as a single sheet of material, significant time and cost savings are possible.

The assay format can be reversed where the patient or organism's DNA is immobilized as the array elements and each array is hybridized with a different mutated allele or genetic marker. The gridded solid support can also be used for parallel non-DNA ELISA assays. Furthermore, the invention allows for the use of all standard detection methods without the need to remove the shallow barrier elements to carry out the detection step.

In addition to the genetic applications listed above, arrays of whole cells, peptides, enzymes, antibodies, antigens, receptors, ligands, phospholipids, polymers, drug cogener preparations or chemical substances can be fabricated by the means described in this invention for large scale screening assays in medical diagnostics, drug discovery, molecular biology, immunology and toxicology.

The multi-cell substrate aspect of the invention allows for the rapid and convenient screening of many DNA probes against many ordered arrays of DNA fragments. This eliminates the need to handle and detect many individual arrays for performing mass screenings for genetic research and diagnostic applications. Numerous microarrays can be fabricated on the same solid support and each microarray reacted with a different DNA probe while the solid support is processed as a single sheet of material.

The following examples illustrate, but in no way are intended to limit, the present invention.

EXAMPLE 1

Genomic-Complexity Hybridization to DNA Microarrays Representing the Yeast *Saccharomyces cerevisiae* Genome with Two-Color Fluorescent Detection

The array elements were randomly amplified PCR (Bohlander, et al., 1992) products using physically mapped lambda clones of *S. cerevisiae* genomic DNA as templates (Riles, et al., 1993). The PCR was performed directly on the lambda phage lysates, resulting in an amplification of both the 35 kb lambda vector and the 5–15 kb yeast insert sequences in the form of a uniform distribution of PCR product between 250–1500 base pairs in length. The PCR product was purified using Sephadex G50 gel filtration (Pharmacia, Piscataway, N.J.) and concentrated by evaporation to dryness at room temperature overnight. Each of the 864 amplified lambda clones was rehydrated in 15 μl of 3xSSC in preparation for spotting onto the glass.

The microarrays were fabricated on microscope slides which were coated with a layer of poly-l-lysine (Sigma). The automated apparatus described in Section III loaded 1 μl of the concentrated lambda clone PCR product in 3xSSC directly from 96 well storage plates into the open capillary printing element and deposited ~5 nl of sample per slide at 380 micron spacing between spots, on each of 40 slides. The process was repeated for all 864 samples and 8 control spots. After the spotting operation was complete, the slides were rehydrated in a humid chamber for 2 hours, baked in a dry 80° vacuum oven for 2 hours, rinsed to remove unabsorbed DNA and then treated with succinic anhydride to reduce non-specific adsorption of the labeled hybridization probe to the poly-l-lysine coated glass surface. Immediately prior to use, the immobilized DNA on the array was denatured in distilled water at 90° for 2 minutes.

For the pooled chromosome experiment, the 16 chromosomes of *Saccharomyces cerevisiae* were separated in a CHEF agarose gel apparatus (Biorad, Richmond, Calif.). The six largest chromosomes were isolated in one gel slice and the ten smallest chromosomes in a second gel slice. The DNA was recovered using a gel extraction kit (Qiagen, Chatsworth, Calif.). The two chromosome pools were randomly amplified in a manner similar to that used for the target lambda clones. Following amplification, 5 micrograms of each of the amplified chromosome pools were separately random-primer labeled using Klenow polymerase (Amersham, Arlington Heights, Ill.) with a lissamine conjugated nucleotide analog (Dupont NEN, Boston, Mass.) for the pool containing the six largest chromosomes, and with a fluorescein conjugated nucleotide analog (BMB) for the pool containing ten smallest chromosomes. The two pools were mixed and concentrated using an ultrafiltration device (Amicon, Danvers, Mass.).

Five micrograms of the hybridization probe consisting of both chromosome pools in 7.5 μl of TE was denatured in a boiling water bath and then snap cooled on ice. 2.5 μl of concentrated hybridization solution (5xSSC and 0.1% SDS) was added and all 10 μl transferred to the array surface, covered with a cover slip, placed in a custom-built single-slide humidity chamber and incubated at 60° for 12 hours. The slides were then rinsed at room temperature in 0.1xSSC and 0.1% SDS for 5 minutes, cover slipped and scanned.

A custom built laser fluorescent scanner was used to detect the two-color hybridization signals from the 1.8x1.8

cm array at 20 micron resolution. The scanned image was gridded and analyzed using custom image analysis software. After correcting for optical crosstalk between the fluorophores due to their overlapping emission spectra, the red and green hybridization values for each clone on the array were correlated to the known physical map position of the clone resulting in a computer-generated color karyotype of the yeast genome.

FIG. 6 shows the hybridization pattern of the two chromosome pools. A red signal indicates that the lambda clone on the array surface contains a cloned genomic DNA segment from one of the six largest yeast chromosomes. A green signal indicates that the lambda clone insert comes from one of the ten smallest yeast chromosomes. Orange signals indicate repetitive sequences which cross hybridized to both chromosome pools. Control spots on the array confirm that the hybridization is specific and reproducible.

The physical map locations of the genomic DNA fragments contained in each of the clones used as array elements have been previously determined by Olson and co-workers (Riles, et al.), allowing for the automatic generation of the color karyotype shown in FIG. 7. The color of a chromosomal section on the karyotype corresponds to the color of the array element containing the clone from that section. The black regions of the karyotype represent false negative dark spots on the array (10%) or regions of the genome not covered by the Olson clone library (90%). Note that the six largest chromosomes are mainly red while the ten smallest chromosomes are mainly green, thus matching the original CHEF gel isolation of the hybridization probe. Areas of the red chromosomes containing green spots and vice-versa are probably due to spurious sample tracking errors in the formation of the original library and in the amplification and spotting procedures.

The yeast genome arrays have also been probed with individual clones or pools of clones that are fluorescently labeled for physical mapping purposes. The hybridization signals of these clones to the array were translated into positions on the physical map of the yeast genome.

EXAMPLE 2

Total cDNA Hybridized to Micro Arrays of cDNA Clones with Two-Color Fluorescent Detection

Twenty-four clones containing cDNA inserts from the plant Arabidopsis were amplified using PCR. Salt was added to the purified PCR products to a final concentration of 3xSSC. The cDNA clones were spotted on poly-l-lysine coated microscope slides in a manner similar to Example 1. Among the cDNA clones was a clone representing a transcription factor HAT4, which had previously been used to create a transgenic line of the plant Arabidopsis, in which this gene is present at ten times the level found in wild-type Arabidopsis (Schena, et al., 1992).

Total poly-A mRNA from wild type Arabidopsis was isolated using standard methods (Maniatis, et al., 1989) and reverse transcribed into total cDNA, using a fluorescein nucleotide analog to label the cDNA product (green fluorescence). A similar procedure was performed with the transgenic line of Arabidopsis where the transcription factor HAT4 was inserted into the genome using standard gene transfer protocols. cDNA copies of mRNA from the transgenic plant are labeled with a lissamine nucleotide analog (red fluorescence). Two micrograms of the cDNA products from each type of plant were pooled together and hybridized to the cDNA clone array in a 10 microliter hybridization reaction in a manner similar to Example 1. Rinsing and detection of hybridization was also performed in a manner similar to Example 1. FIG. 8 shows the resulting hybridization pattern of the array.

Genes equally expressed in wild type and the transgenic Arabidopsis appeared yellow due to equal contributions of the green and red fluorescence to the final signal. The dots are different intensities of yellow indicating various levels of gene expression. The cDNA clone representing the transcription factor HAT4, expressed in the transgenic line of Arabidopsis but not detectably expressed in wild type Arabidopsis, appears as a red dot (with the arrow pointing to it), indicating the preferential expression of the transcription factor in the red-labeled transgenic Arabidopsis and the relative lack of expression of the transcription factor in the green-labeled wild type Arabidopsis.

An advantage of the microarray hybridization format for gene expression studies is the high partial concentration of each cDNA species achievable in the 10 microliter hybridization reaction. This high partial concentration allows for detection of rare transcripts without the need for PCR amplification of the hybridization probe which may bias the true genetic representation of each discrete cDNA species.

Gene expression studies such as these can be used for genomics research to discover which genes are expressed in which cell types, disease states, development states or environmental conditions. Gene expression studies can also be used for diagnosis of disease by empirically correlating gene expression patterns to disease states.

EXAMPLE 3

Multiplexed Colorimetric Hybridization on a Gridded Solid Support

A sheet of plastic-backed nitrocellulose was gridded with barrier elements made from silicone rubber according to the description in Section IV-A. The sheet was soaked in 10xSSC and allowed to dry. As shown in FIG. 12, 192 M13 clones, each with a different yeast inserts were arrayed 400 microns apart in four quadrants of the solid support using the automated device described in Section III. The bottom left quadrant served as a negative control for hybridization, while each of the other three quadrants was hybridized simultaneously with a different oligonucleotide using the open-face hybridization technology described in Section IV-A. The first two and last four elements of each array are positive controls for the calorimetric detection step.

The oligonucleotides were labeled with fluorescein, which was detected using an anti-fluorescein antibody conjugated to alkaline phosphatase that precipitated an NBT/BCIP dye on the solid support (Amersham). Perfect matches between the labeled oligos and the M13 clones resulted in dark spots visible to the naked eye and detected using an optical scanner (HP ScanJet II) attached to a personal computer. The hybridization patterns are different in every quadrant indicating that each oligo found several unique M13 clones from among the 192 with a perfect sequence match. Note that the open capillary printing tip leaves detectable dimples on the nitrocellulose which can be used to automatically align and analyze the images.

Although the invention has been described with respect to specific embodiments and methods, it will be clear that various changes and modifications may be made without departing from the invention.

We claim:

1. A method of forming a microarray of discrete analyte-assay regions on a solid support, where each discrete region in the microarray has a selected, analyte-specific reagent, said method comprising,

(a) loading an aqueous solution of a selected analyte-specific reagent in a reagent-dispensing device having an elongate capillary channel adapted to hold a quantity of the reagent solution and having a tip region at which the solution in the channel forms a meniscus,

(b) tapping the tip of the dispensing device against a solid support at a defined position on the surface, with an impulse effective to break the meniscus in the capillary channel and deposit a selected volume between 0.002 and 2 nl of solution on the surface, and

(c) repeating steps (a) and (b) until said microarray is formed.

2. The method of claim 1, wherein the reagents used to form the discrete regions in the microarray are distinct nucleic acid strands and wherein steps (a) and (b) are repeated until the microarray has about 100 or more discrete regions of distinct nucleic acid strands per cm^2 of solid support.

3. The method of claim 1, wherein the reagents used to form the discrete regions in the microarray are distinct nucleic acid strands and wherein steps (a) and (b) are repeated until the microarray has about 1000 or more discrete regions of distinct nucleic acid strands per cm^2 of solid support.

4. The method of claim 2, wherein the channel is open-sided.

5. The method of claim 3, wherein the channel is open-sided.

6. The method of claim 4, wherein the volume is between 0.002 and 0.25 nl.

7. The method of claim 5, wherein the volume is between 0.002 and 0.25 nl.

* * * * *

E

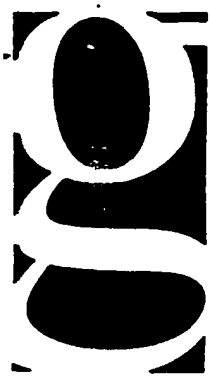
nature genetics
volume 14 no. 4 december 1996

DNA chips, diagnostics and genomics

Rieger syndrome

QTLs and epistasis

A BRCA1-binding protein



Editorial Office
545 National Press Building
Washington DC 20045
Tel: (202) 626-2513
Fax: (202) 626-0970
email: natgen@naturedc.com

WWW: genetics.nature.com

Cover art: Ken Krattenmaker

ature Genetics (ISSN 1061-4036) is published monthly by Nature Publishing Co., headquartered at 345 Park Avenue South, New York, NY 10010, which is owned by Nature America c., a subsidiary of Macmillan Magazines Ltd. of London. Editorial Office: Nature Genetics, 545 National Press Building, Washington DC 20045. Telephone (202) 626 2513, Fax (202) 6 0970. e-mail natgen@naturedc.com. North American Advertising: Nature Genetics, 345 Park Avenue South, New York, NY 10010. Telephone (212) 726 9200, Fax (212) 696 9606. uropean Advertising: Nature Genetics, Porters South, Crinan Street, London N1 9XW, UK. Telephone (0171) 833 4000, Fax (0171) 843 4596. New subscriptions, renewals, anges of address, back issues and all customer service questions should be addressed to: Nature Genetics Subscription Dept, 345 Park Avenue South, New York, NY 10010, SA. Telephone (212) 726 9200, Fax (212) 696 9606. Outside North America: Nature Genetics, Macmillan Magazines Ltd, Porters South, Crinan Street, London N1 9XW, UK. lephone (0171) 833 4000, Fax (0171) 843 4596. Annual subscription rates: US/Canada: US$495 (institutional/corporate), US$195 (individual making personal payment), Canada id 7% for GST, BN: 14091 1595 RT; UK: £350 (institutional/corporate): £175 (individual making personal payment). Japan: Contact Nature Japan, K.K., Shin-Mitsuke Bldg. 4F, 3-6 higaya Tamachi, Shinjuku-ku, Tokyo 162. Reprints: Nature Genetics Reprints Dept, 345 Park Avenue South, New York, NY 10010-1707. Periodicals

postmaster: Send address changes to Nature Genetics Subscription Dept, 345 Park Avenue South, New York, NY 10010. Executive Officers of Nature America, Inc.: Nicholas Byam law, Chairman of the Board; Mary Waltham, President; Edward Vaks, Secretary-Treasurer. Published in Japan by Nature Japan KK, Shin-Mitsuke Building, 3F, 3-6 Ichigaya Tamachi, injuku-ku, Tokyo 162, Japan. Printed in the USA by Cadmus Journal Services.

Nature Publishing Co.
345 Park Avenue South
10th floor
New York, NY 10010-1707
Tel: (212) 726-9200
Fax: (212) 696-9606

President–Publisher
Mary Waltham

Vice President Sales
Marion Delaney

Vice President Marketing
James A. Skowrenski

American Advertising Sales Manager
Sande T. Giaccone (New York)

European Advertising Sales Manager
Kathryn Wayman (London)

Classified Advertising Sales Manager
Erika A. Simon (New York)
Mike Grant (London)

Assistant Classified Sales Manager
Benjamin Crowe (New York)

Production & Information Systems Director
Nick Kemp

Circulation Manager
Moira Musto (New York)
Vic Harman (London)

Group Marketing Manager
Gina Dzurenda

Macmillan Magazines Ltd
Porters South
Crinan Street
London N1 9XW
Tel: 44 (0)171 833 4000
Fax: 44 (0)171 843 4596

Managing Director
Andy Barker

Publishing Director
Judy Sutherland

Editor-in-Chief,
Nature publications
Philip Campbell

Publishing Director
Steve Walker

Nature Japan KK
Yamatomitsuke Bldg
Kagurazaka Tamachi
Shinjuku-ku
Tokyo 162
Phone 03 3267 8751
03 3267 8746

Publisher
Neil Swinbanks

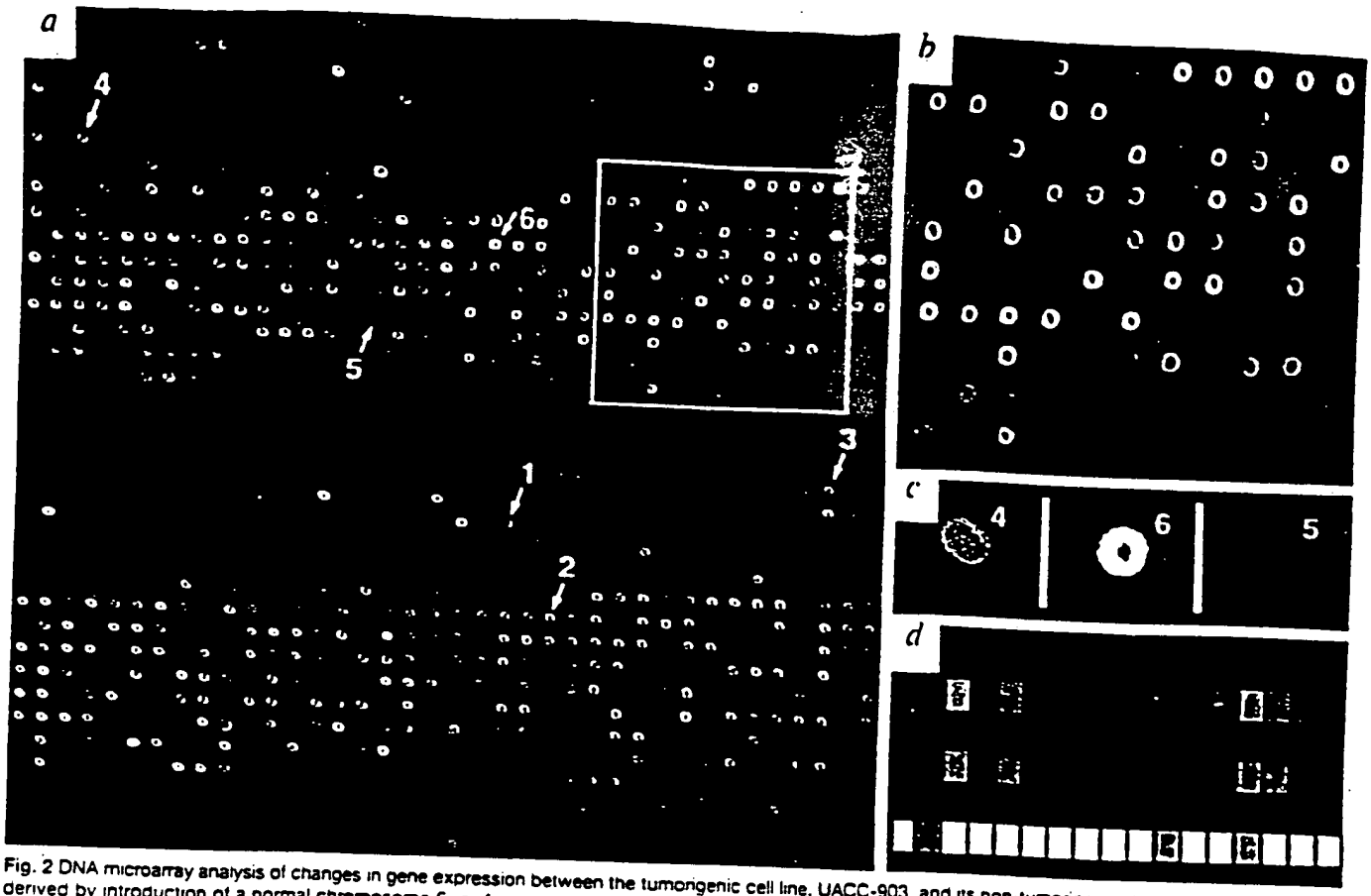


Fig. 2 DNA microarray analysis of changes in gene expression between the tumorigenic cell line, UACC-903, and its non-tumorigenic derivative, UACC903(+6), derived by introduction of a normal chromosome 6. *a*, A ratio image of the results of simultaneous hybridization of Rhodamine110-labelled cDNA (green) from UACC-903 and Cy3-labelled cDNA (orange-red) from UACC-903(+6) to a microarray. To produce this image, the scan images corresponding to each fluorescent probe were combined as the appropriate colour channels in a single image. Arrows indicate the location within the array of the corresponding genes analysed by northern blotting (Fig. 3). *b*, A magnified image of the area of the array boxed in white in (*a*). *c*, Magnified image of three cDNAs indentified by arrows in (*a*), representing the cDNAs for: left, *MCAF/MCP-1* (r/g ratio >10); centre, *β-actin* (r/g ratio 1.04); and right, *α-1-antichymotrypsin* (r/g ratio 0.2) (see Fig. 3). *d*, simplified representation of ratio hybridization results. Quantitative fluorescence intensity data is extracted from each array target. The average target colour ratio determines the hue of each box and the average intensity determines the brightness of each box. In this image, the order of the boxes corresponds to their original order in the microtiter plate from which they were printed. Duplicate printings of the same plate can be examined side by side, as in the first two rows shown here, to assess reproducibility of the hybridization results (see text). Numbered arrows indicate the location within the array corresponding to genes analysed by northern blotting in Fig. 3.

Methods

Generation of microarrays, hybridization, scanning. The preparation of coated microscope slides and subsequent robotic printing of DNA was carried out in a manner similar to that described[7]. Briefly, pre-cleaned glass slides were treated with poly-L-lysine solution (Sigma) to form an adhesive surface for printing. PCR products, purified by ethanol purification, were resuspended in 3x SSC. A custom built arraying robot picked up and deposited small volumes (~5 nanoliters) of DNA onto the slides. After printing, the slides were washed in a 0.2% SDS solution. The remaining bound DNA was denatured by submerging the slides in 95 °C distilled water for 2 min followed by a brief wash with 95% ethanol. DNA was UV crosslinked to the slides (Stratagene Stratalinker, 60 mJ). To prevent non-specific probe binding, the slides were blocked by rinsing in a solution of 70 mM succinic anhydride dissolved in 0.1 M boric acid pH 8.0, containing 35% 1-methyl-2-pyrrolidinone (Aldrich). Additional protocols and parts list pertaining to microarray fabrication can be obtained from http://cmgm.stanford.edu/pbrown.

Purified, labelled cDNA was resuspended in 11 µl of 3.5× SSC containing 4 µg of poly (dA)$^{+}$ DNA, 2.5 µg *E. coli* tRNA, 4 µg of human Cot1 DNA (Gibco BRL), and 0.3 µl of 10% SDS. Prior to hybridization, the solution was boiled for 2 min then allowed to cool to room temperature. Hybridization was carried out at 62 °C for ~14 h in a water bath. Prior to scanning, slides were washed in 2× SSC, 0.2% SDS for 5 min and 0.2× SSC for 1 min [illegible] local laser microscope built by S. Smith with software written by N. Zav. A separate scan, using the appropriate excitation line, was done for each of the two fluorophores used. Data was collected at a maximum resolution of 9 microns/pixel with 12 bits of depth.

Probe preparation and labelling. RNA was extracted from cells using the Trizol reagent (LTI Inc.), following the manufacturer's directions. cDNA probes were synthesized from singly oligo dT-selected (Pharmacia) mRNA pools. Fluorescently labelled cDNA was prepared from mRNA by oligo dT-primed polymerization using SuperScript II reverse transcriptase (LTI Inc.). The pool of nucleotides in the labelling reaction was 0.5 mM dGTP, dATP and dCTP and 0.2 mM dTTP. Fluorescent nucleotides, Rhodamine 110 dUTP (Perkin Elmer Cetus) or Cy3 dUTP (Amersham), were present at 0.1 mM. Probes were purified by gel chromatography (BioSpin 6/BioRad) and ethanol precipitation.

Selection of cDNA elements and generation of control templates. Synthetic cDNAs were prepared by cloning random *Bam*HI and *Hin*dIII ended fragments of *E. coli* DNA in the vector pSP64 poly (A)$^{+}$ (Promega), linearizing isolated plasmid DNA with *Eco*RI and synthesizing poly (A)$^{+}$ tailed RNA complementary to the insert from the resident SP6 promoter (Promega). Prior to use, the synthesized RNAs were selected on oligo dT cellulose. The largest group of cDNAs consisted of 674 cDNA clones from the LNIB arrayed normalized infant brain [illegible]. These [illegible] were selected to [illegible] every NIB library member that corresponded to a named gene according

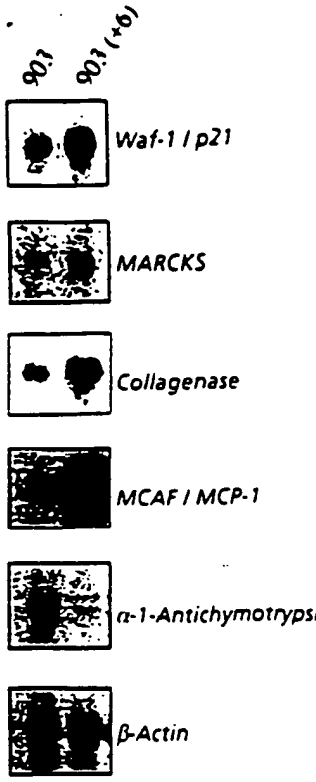


Fig. 3 Northern hybridization substantiating the consistency of the cDNA microarray results. Corresponding locations within the cDNA microarray illustrated in Fig. 2*a* are provided for 1) *Waf-1/p21*; 2) *MARCKS*; 3) *collagenase*; 4) *MCAF/MCP-1*; 5) *α-1-antichymotrypsin*; and 6) *β-actin*. The signal detected by a radio-labelled *β-actin* probe represents a control for loading variance, with a red/green ratio observed on the cDNA microarray (Fig. 2*a,c*) for *β-actin* of 1.04.

to the UniGene EST clustering system[21,22]. The second largest group of clones consisted of 183 sequenced cDNA clones generated by subtraction of cDNA from the chromosome-6 suppressed non-tumorigenic UACC-903 (+6) cell line with cDNA from its parental tumorigenic cell line UACC-903 (ref. 9). Approximately 100 additional genes (total 870 genes arrayed) were obtained from EST libraries on the basis of their expression pattern (tissue specific, and so on). Each array included the following hybridization controls: plasmid vector, lambda, ϕX174 phage, total human DNA, human Cot1 DNA, and poly (A)$^+$. The synthetic standards used for normalization of signals in each wavelength were also arrayed. Controls were included in each quadrant of the array to assess the reproducibility of the hybridization signal. Two plates of cDNA clones (derived from the UACC-903 subtracted library) were also arrayed in duplicate. Fidelity of the Unigene array relative to dbEST was tested by sequencing of a random sample of 11 clones used for microarray construction. All sequences were identical with the corresponding dbEST entries. Additionally, each microarrayed cDNA from the UACC-903 subtracted library was sequenced. A listing of cDNAs comprising this microarray which were derived from the Unigene and 'housekeeping' panel can be obtained from http://www.nih.gov/DIR/LCG/ARRAY/expn.html.

Northern blot analysis. Total RNA, 10 μg per lane, was electrophoresed in 1.2% agarose-formaldehyde gels and transferred onto nylon membrane (Hybond-N$^+$, Amersham) by capillary blotting overnight. For DNA probes insert fragments from the Soares IMAGE cDNA library[10] were obtained by vector PCR for p21, MARCKS, α-1-antichymotrypsin and β-actin. Probes for fibroblast collagenase and MCAF/MCP-1 were isolated from a UACC-903(+6) enriched cDNA library[9] with all probes labelled by random priming. Filters were washed to a stringency of 0.1× SSC at 42 °C for 20 min.

Web sites. http://cmgm.stanford.edu/pbrown for protocols and parts list pertaining to microarray fabrication. http://www.nchgr.nih.gov/DIR/LCG/ARRAY/expn.html for a listing of cDNAs comprising this microarray which were derived from the Unigene and 'housekeeping' panel.

Acknowledgements

Work in P.O.B.'s laboratory is supported in part by the Howard Hughes Medical Institute and National Center for Human Genome Research (HG00450). We would like to acknowledge the excellent technical and graphic assistance of X. He, T. Hofmann, Y. Jiang, J. Leuders, D. Leja and B. Walker. J.D. was supported by NIH grant 2T32RM07276-21. P.O.B. is an assistant investigator of the Howard Hughes Medical Institute.

Received 15 October; accepted 8 November, 1996.

1. Vogelstein, B. & Kinzler, K.W. The multistep nature of cancer. *Trends Genet.* 9, 138–141 (1993).
2. Weinberg, R.A. The molecular basis of oncogenes and tumor suppressor genes. *Ann. NY Acad. Sci.* 758, 331–338 (1995).
3. Levine, A.J. The tumor suppressor genes. *Annu. Rev. Biochem.* 62, 623–651 (1993).
4. Trent, J.M. *et al.* Tumorigenicity in human melanoma cell lines controlled by introduction of human chromosome 6. *Science* 247, 568–571 (1990).
5. Su, Y. *et al.* Reversion of monochromosome-mediated suppression of tumorigenicity in malignant melanoma by retroviral transduction. *Cancer Res.* 56, 3186–3191 (1996).
6. Schena, M., Shalon, D., Davis, R.W., & Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470 (1995).
7. Shalon, D., Smith, S.J. & Brown, P.O. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.* 6, 639–645 (1996).
8. Schena, M. *et al.* Parallel human genome analysis: microarray-based expression of 1000 genes. *Proc. Natl. Acad. Sci. USA* 93,10539–11286 (1996).
9. Ray, M.E., Su, Y.A., Meltzer, P.S. & Trent, J.M. Isolation and characterization of genes associated with chromosome-6 mediated tumor suppression in human malignant melanoma. *Oncogene* 12, 2527–2533 (1996).
10. Soares, M.B. *et al.* Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. Sci. USA* 91, 9228–9232 (1994).
11. Boguski, M.S. & Schuler, G.D. ESTablishing a human transcript map *Nature Genet.* 10, 369–371 (1995).
12. Vijayasaradhi, S., Doskoch, P.M., Wolchok, J. & Houghton, A.N. Melanocyte differentiation marker gp75, the brown locus protein, can be regulated independently of tyrosinase and pigmentation. *J. Invest. Dermatol.* 105, 113–119 (1995).
13. Vijayasaradhi, S., Xu, Y., Bouchard, B. & Houghton, A.N. Intracellular sorting and targeting of melanosomal membrane proteins: identification of signals for sorting of the human brown locus protein, gp 75. *J. Invest. Dermatol.* 130, 807–820 (1995).
14. Nakao, J. *et al.* Expression of proteolipid protein gene is directly associated with secretion of a factor influencing oligodendrocyte development. *J. Neurochem.* 6455, 2396–2403 (1995).
15. Graves, D.T., Barnhill, R., Galanopoulos, T. & Antoniades, H.N. Expression of monocyte chemotactic protein-1 in human melanoma in vivo. *Am. J. Pathol.* 140, 9–14 (1992).
16. Kristensen, M.S., Deleuran, B.W., Larsen, C.G., Thestrup-Pedersen, K. & Paludan, K. Expression of monocyte chemotactic and activating factor (MCAF) in skin related cells. A comparative study. *Cytokine* 5, 520–524 (1993).
17. Huang, S., Xie, K. Singh, R.K., Gutman, M. & Bar-Eli, M. Suppression of tumor growth and metastasis of murine renal adenocarcinoma by syngeneic fibroblasts genetically engineered to secrete the JE/MCP-1 cytokine. *J. Interferon Cytokine Res.* 15, 655–665 (1995).
18. El-Deiry, W.S. *et al.* WAF1, a potential mediator of p53 tumor suppression. *Cell* 75, 817–825 (1993).
19. Miele, M.E. *et al.* Metastasis suppressed, but tumorigenicity and local invasiveness unaffected, in the human melanoma cell line MeJuSo after introduction of human chromosomes 1 or 6. *Mol. Carcinog.* 15, 284–299 (1996).
20. Jiang, H. *et al.* The melanoma differentiation-associated gene mda-6, which encodes the cyclin-dependent kinase inhibitor p21, is differentially expressed during growth, differentiation and progression in human melanoma cells. *Oncogene* 10, 1855–1864 (1995).
21. Schuler, G.D. *et al.* A gene map of the human genome. *Science* 274, 540–546 (1996).
22. Lennon, G., Auffray, C., Polymeropoulos, M. & Soares, M.B. The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics* 33, 151–152 (1996).

ISSN 1088-9051

GENOME RESEARCH

July 1996 Volume 6 Number 7

INCLUDING PCR

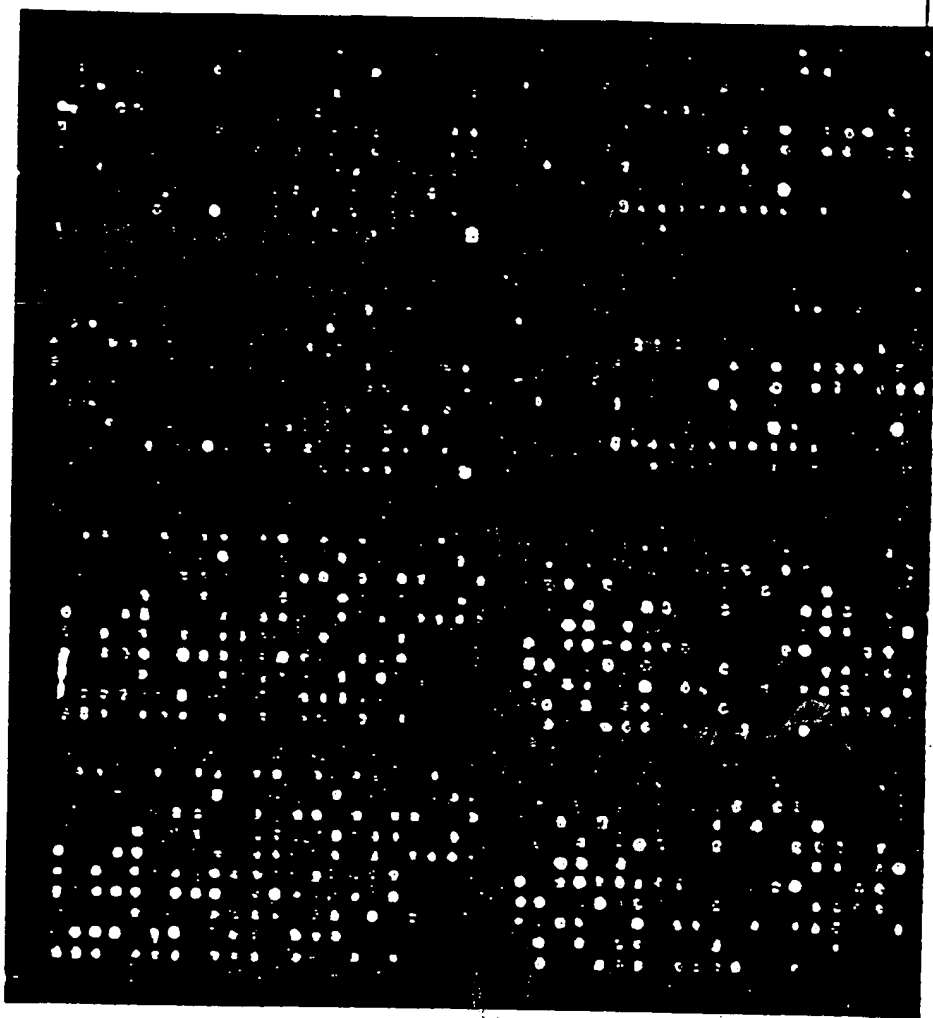
IBD Mapping in Livestock

Sequence of 500-kb *Rhizobium* Replicon

Human Y Chromosome Haplotypes

BAC Mapping of Extrachromosomal Structure

DNA Microarray System



Cold Spring Harbor Laboratory Press

F

A DNA Microarray System for Analyzing Complex DNA Samples Using Two-color Fluorescent Probe Hybridization

Dari Shalon,[1,4] Stephen J. Smith,[3] and Patrick O. Brown[1,2,5]

[1]Howard Hughes Medical Institute and Departments of [2]Biochemistry and [3]Molecular and Cellular Physiology, Stanford University, Stanford, California 94305

Detecting and determining the relative abundance of diverse individual sequences in complex DNA samples is a recurring experimental challenge in analyzing genomes. We describe a general experimental approach to this problem, using microscopic arrays of DNA fragments on glass substrates for differential hybridization analysis of fluorescently labeled DNA samples. To test the system, 864 physically mapped λ clones of yeast genomic DNA, together representing >75% of the yeast genome, were arranged into 1.8-cm × 1.8-cm arrays, each containing a total of 1744 elements. The microarrays were characterized by simultaneous hybridization of two different sets of isolated yeast chromosomes labeled with two different fluorophores. A laser fluorescent scanner was used to detect the hybridization signals from the two fluorophores. The results demonstrate the utility of DNA microarrays in the analysis of complex DNA samples. This system should find numerous applications in genome-wide genetic mapping, physical mapping, and gene expression studies.

Many problems in genome analysis depend on determining what specific sequences are represented in a complex DNA or RNA sample and at what abundance, for example, what genes are represented in a specific chromosome band or YAC clone, what intervals are amplified or deleted in a particular cancer cell, or what genes are expressed in specific cells under specific conditions. As a general approach to this problem, we have developed a system for making microarrays of DNA samples on glass substrates, probing them by hybridization with complex fluorescent-labeled probes, and using a laser-scanning microscope to detect the fluorescent signals representing hybridization. Fluorescent labeling allows for simultaneous hybridization and separate detection of the hybridization signal from two or more probes. This in turn allows very accurate and reliable measurement of the relative abundance of specific sequences in two complex samples.

RESULTS

Array Hybridization Pattern

Figure 1 shows the two-color fluorescent scan of a yeast genomic array following hybridization with a mixed probe consisting of lissamine-labeled DNA from the 6 largest yeast chromosomes together with fluorescein-labeled DNA from the 10 smallest yeast chromosomes. A red color indicates that yeast sequences present in the lissamine-labeled hybridization probe hybridized to an array element. A yellow-green color indicates that yeast sequences present in the fluorescein-labeled hybridization probe hybridized to an array element. An orange color indicates cross-hybridization of both chromosome pools to an array element (e.g., dispersed repetitive elements, such as Ty1 elements).

Each clone was spotted twice, resulting in duplicate hybridization patterns in adjacent quadrants of the array. Control DNA spots, which were randomly amplified in the same manner as the λ clone array elements, are located in the bottom corner of each quadrant. "A" points to a pair of spots containing total yeast genomic DNA. These spots appear orange because both chromosome pools hybridized to yeast genomic DNA. The negative controls are as follows: "B" points to a pair of spots of wild-type λ DNA, "C" points to a pair of human genomic DNA spots, and "D" points to a pair of ϕX174 DNA spots. The lack of a hybridization signal at these three negative control spots indicates that the hybridization was specific for yeast sequences.

[4]Present address: Synteni, Inc., Palo Alto, California 94305.
[5]Corresponding author.
E-MAIL pbrown@cmgm.stanford.edu; http://cmgm.stanford.edu/pbrown; FAX (415) 723-1399.

6:639–645 ©1996 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/96 $5.00

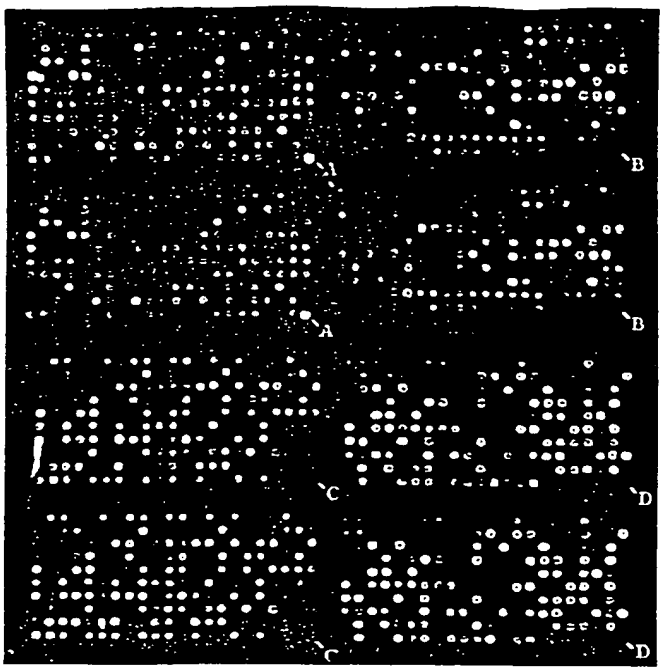


Figure 1 Two-color fluorescent scan of a 1.8-cm × 1.8-cm yeast array of λ clones of yeast genomic DNA. The DNA spots are spaced at a distance of 380 μm from center to center. A probe mixture consisting of DNA from the 6 largest yeast chromosomes (4, 7, 12, 13, 15, 16) labeled with lissamine (red dots) and DNA from the 10 smallest yeast chromosomes (1, 2, 3, 5, 6, 8, 9, 10, 11, 14) labeled with fluorescein (yellow-green dots) was hybridized to the array. A pair of yeast genomic DNA spots (*A*) served as a positive control. The three negative controls are λ DNA (*B*), human genomic DNA (*C*), and φX174 DNA (*D*).

Karyotype Depiction of the Array Hybridization Pattern

The inserts contained in the arrayed λ clones have been mapped physically (Riles et al. 1993). The clones are arrayed in a random but known order on the array. Therefore, using the identity of each clone along with its physical map information, the pattern of hybridization to the yeast array can be represented in the form of a karyotype of the yeast genome, as shown in Figure 2. The color of any segment of the ideogram representing an individual chromosome on the karyotype is directly determined by the ratio of red and green hybridization signals at the array positions of the corresponding clones. The lengths of the discrete colored segments of each chromosome correspond to the physical lengths of the yeast inserts. The chromosome segments colored black represent either intervals of the genome that are not represented by clones in the library (90%) or false-negative hybridization signals on the array (10%). Most of these false negatives are attributable to failures of the PCR amplification of the λ clones, though occasional failures of the arraying process or nonuniform surface preparation could account for a small fraction of the false-negative signals. The large gap on chromosome 12 is the region coding for ribosomal DNA that was not represented among the arrayed clones. Genomic intervals represented by overlapping clones were assigned a color based on the hybridization signals of only one of the overlapping clones, chosen at random.

Note that in this representation of a yeast karyotype, the largest six chromosomes are mainly colored red. This indicates that most of the arrayed clones that were mapped previously to these six large chromosomes hybridized primarily to the lissamine-labeled probe prepared from the corresponding purified chromosomes. Conversely, the smallest 10 chromosomes are mainly colored green in this image, matching the original CHEF gel isolation of the chromosomes used as the hybridization probe. The experiment was repeated with the yeast genome split into six discrete chromosome pools containing 2–4 chromosomes per pool using CHEF gel electrophoresis. The chromosomes in each pool were extracted from the gel, amplified, and fluorescently labeled. The six chromosome pools were hybridized to six separate yeast arrays. Forty-four λ clones gave a positive hybridization signal on all six arrays indicating that they contain yeast repetitive sequences (data not shown). These 44 clones and 10 clones with very weak hybridization signals were not included in the data set used to produce this karyotype.

There were ~40 anomalous clones, which appear in this karyotype representation as green bands on the otherwise red chromosomes or red

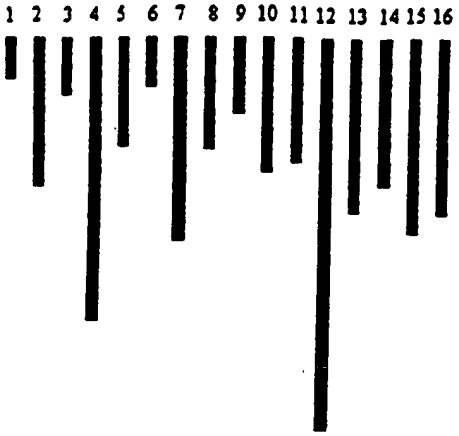


Figure 2 Computer-generated ideogram representing a karyotype of *S. cerevisiae*, based on the normalized hybridization signals from the array shown in Fig. 1. Note that the 6 largest chromosomes are mainly red and the 10 smallest chromosomes are mainly green. Black stripes represent intervals not represented by clones in the array or for which the corresponding clones gave false-negative hybridization signals.

bands on the otherwise green chromosomes. Four randomly chosen examples of these anomalous clones were analyzed by hybridizing the clones to vertical strips cut from a Southern blot of CHEF gel-separated yeast chromosomes. In each case, the hybridization patterns of the anomalous clones corroborated the chromosomal locations assigned by the microarray hybridization results (data not shown). Two clones that were thought to map to the 10 smallest chromosomes were found to hybridize preferentially to the probe representing the 6 largest chromosomes and thus appear as anomalous red bands on the karyotype. Both hybridized to one of the six largest chromosomes on the Southern blot. Similarly, two clones that appear as anomalous green bands on the karyotype were found to hybridize to one of the 10 smallest chromosomes on the Southern blot. Thus, the anomalous clones are probably the result of sample tracking errors or, possibly, of errors in the published restriction-digest-based physical map on which the karyotype representation was based (Riles et al. 1993).

DISCUSSION

The DNA microarray hybridization system reported here is conceptually and functionally similar to fluorescent in situ hybridization (FISH) to metaphase chromosomes, with three important differences. First, the target elements of the microarrays can, in principle, be any length or composition, from megabase YAC clones or microdissected chromosome bands to individual cDNA clones, to short oligonucleotides. This versatility allows the user to choose characteristics, such as the mapping resolution and genetic complexity of each array element, to suit a particular application. Second, the hybridization signals are localized to discrete elements of known size and location, making them easier to identify and quantitate than the hybridization signals from irregularly shaped metaphase spreads. Third, microarrays are more consistent and potentially amenable to automated production, hybridization, and data analysis than metaphase spreads.

Arrays of DNA samples on porous membranes, for example, dot blots, have long been used as a basic tool in molecular biology. Dot-blot membranes are usually at least 8 × 12 cm in size, require the use of milliliter volumes of hybridization solution, and are limited, owing to autofluorescence and scattering, to radioactive, chemiluminescent, and colorimetric hybridization detection methods (Ross et al. 1992). Microarrays made on glass surfaces, on the other hand, can be mass-produced and are comparatively inexpensive, convenient, and compatible with fluorescent hybridization detection methods. Furthermore, a glass surface, when appropriately treated, has very low nonspecific binding of labeled hybridization probes, resulting in lower backgrounds than are encountered typically with porous membranes. For hybridizations with very complex probes, the concentration of the labeled probe DNA is a limiting factor in the sensitivity of the assay. Minimizing the volume of the probe solution in a hybridization, by restricting the target to a small area and by using a nonporous substrate, makes it practical to achieve very high probe concentrations.

One important advantage of fluorescently labeled probes is that, unlike most radioactive and chemiluminescent signals, fluorescent signals do not disperse and therefore allow for very dense array spacing. A unique, and probably the most important, advantage of fluorescent probes is that the hybridization signals from two or more differently labeled probes hybridized to the same target element can be detected separately. In this way, two-color hybridization detection allows for a direct and quantitative comparison of the

abundance of specific sequences between two probe mixtures that are hybridized competitively to a single array. The absolute intensity of a hybridization signal at a particular element in an array can vary owing to experimental factors such as variations in the amount of DNA deposited on the array, variations in the hybridization or wash conditions between experiments, or variations in the hybridization characteristics of the different DNA sequences on the array. The ratio of the two signals at any element in an array, however, is relatively insensitive to these confounding factors because they affect both probe mixtures equivalently. This ratio therefore accurately reflects the relative abundance of the cognate sequence in the two probe samples. This is the principle underlying the technique of comparative genomic hybridization (CGH), which is used to detect changes in the copy number of specific chromosomes or chromosomal regions (Kallioniemi et al. 1992). CGH is based on measuring the relative fluorescent hybridization intensities of two genomic-complexity hybridization probes, for example, probes representing genomic DNA from normal and affected tissue samples, which are labeled with two distinct fluorophores and hybridized simultaneously to a metaphase spread. DNA microarray representations of the human genome may provide a more convenient and higher resolution alternative to metaphase chromosomes for CGH.

Cross-hybridization between related sequences is an important problem faced by any hybridization-based assay, including the DNA microarray assay described here. Studies are now in progress to quantitate the extent of cross-hybridization between related sequences of varying homology and length, in DNA microarray hybridizations. The stringency of hybridization and washing can be controlled by varying the salt concentration and temperature as in conventional membrane-based hybridizations. Cross-hybridization caused by repetitive sequences can be minimized by prehybridization of the probe or array with vast excess of unlabeled copies of the repetitive sequences.

Alternative methods have been described for making microarrays of very short DNA sequences, involving photolithography (Pease et al. 1994) or physical masking (Maskos and Southern 1992) methods. These in situ synthesis methods are inherently limited to low complexity array elements consisting of oligonucleotides. For complex-probe hybridizations, the specificity of hybridization is improved by using DNA fragments substantially longer than oligonucleotides. Moreover, the in situ synthesis approaches to array fabrication depend on prior knowledge of the sequence to be recognized by each array element. The approach described here makes microarrays by transferring tiny volumes of DNA samples from microwell storage plates to a solid substrate. Thus, nucleic acids (or other molecules) of virtually any length or any origin can be arrayed, and knowledge of their sequences is not required.

The arrays used in these experiments do not represent the maximal achievable density of elements. We have found that the spacing between the spots can be decreased by shrinking the contact area of the printing tip and by increasing the hydrophobicity of the glass surface. Microarrays with 100-μm feature size have been tested successfully in pilot experiments (data not shown). Assuming the projected availability of the appropriate physically mapped human genomic clones (Hudson et al. 1995), arrays at 100-μm spacing would allow for 10,000 discrete intervals of the human genome to be represented in a 1-cm^2 array. Such an array could be used for mapping at a resolution of <0.5 Mb. Experiments are in progress to explore the feasibility of such arrays.

Our initial motivation for developing these microarrays arose from the need for abundant and inexpensive genomic arrays for genomic mismatch scanning (GMS) (Nelson et al. 1993), a method of genetic linkage analysis based on identification of the regions of "identity by descent" between affected relative pairs using a single complex-probe hybridization to an array of genomic clones. Experiments using these arrays to map quantitative trait loci in yeast by GMS are currently in progress (J. deRisi, D. Lashkari, L. Penland, L. McAllister, J. McCusker, R. Davis, and P.O. Brown, unpubl.).

Microarrays of cDNA clones, prepared using the system described here, have been used for quantitative monitoring of gene expression patterns in *Arabidopsis* (Schena et al. 1995), *S. cerevisiae* (D. Lashkari, J. deRisi, L. Penland, P.O. Brown, and R. Davis, unpubl.), and human tissues (J. deRisi, M. Bittner, P. Meltzer, L. Penland, J. Trent, and and P.O. Brown, unpubl.). We anticipate that DNA microarrays of the kind described here will be useful in additional applications for which conventional dot blots, high-density gridded arrays on porous membranes, or FISH are currently used. These potential applica-

tions include comparative genomic hybridization (Kallioniemi et al. 1992), sequencing by hybridization (Drmanac et al. 1993), physical mapping of cloned or amplified sequences (Billings et al. 1991), and economical distribution of reagents for integrated genetic and physical mapping based on a common set of arrayed clones (Zehetner and Lehrach 1994).

METHODS

Amplification of Target DNA Elements

The array elements were prepared from physically mapped λ clones (Riles et al. 1993). The λ clones were amplified using randomly primed polymerase chain reaction (PCR) based on published and unpublished protocols (Bohlander et al. 1992; S. Nelson, unpubl.). The phage lysates were amplified in a 10-μl PCR reaction using 5 μM final concentration of primer A (GCTATCTTCAAGATCANNNNNN), 200 μM dNTPs, and 1 unit of *Taq* polymerase. Round A consisted of five cycles at 94°C for 1 min, 25°C for 1.5 min, 25–72°C over 7 min, and 72°C for 3 min using *Taq* polymerase (BMB). For round B, the reaction volume was brought up to 100 μl for a final concentration of 2 μM of primer B (GCTATCTTCAAGATCA), 200 μM dNTPs, and 4 units of *Taq* polymerase. Round B consisted of 30 cycles of 94°C for 1 min, 56°C for 2 min, and 72°C for 3 min. The amplification was performed in 96-well plates using crude phage lysates as the templates, resulting in an amplification of both the 35-kb λ vector and the 5-kb to 15-kb yeast insert sequences as a distribution of PCR products between 250 bp and 1500 bp in length.

The PCR products were purified and transferred into TE (10 mM Tris, 1 mM EDTA at pH 8.0) buffer using Sephadex G50 gel filtration (Pharmacia) and evaporated to dryness at room temperature overnight. Each of the 864 amplified λ clones was rehydrated in 15 μl of 3× SSC (20× SSC = 3 M NaCl, 0.3 M Na_3 citrate) in preparation for spotting onto the glass under normal room temperature conditions.

Preparation of DNA Microarrays

The microarrays were fabricated on poly-L-lysine coated microscope slides (Sigma). A custom-built arraying machine, consisting of four tweezer-like printing tips mounted 9 mm apart on a computer-controlled robotic stage (Shalon 1996), loaded 1 μl of the concentrated PCR product directly from corresponding clusters of four wells of 96-well storage plates and deposited ~5 nl of each sample onto each of 40 slides. Surface tension loaded the sample into the printing tip directly from the microwell plate and held the sample in the tip during the printing operation. Printing was achieved by lightly tapping the tip against the glass surface. The open-capillary design allowed for rapid rinsing and drying of the tips between samples. Figure 3 shows the layout of the arraying machine. Figure 4 shows a detailed view of the four printing tips and the staggered printing pattern on the microscope slides. Adjacent samples were spotted 380 μm apart on the slides. After each set of four samples was printed onto 40 slides, the printing tips were rinsed with a jet of water for 2 sec and then dried by lowering the tips onto a sponge for 2 sec. The process was repeated for all 864 samples and eight control spots.

After the sporting operation was complete, the slides were rehydrated in a humid chamber at room temperature for 2 hr, baked in an 80°C vacuum oven for 2 hr, then rinsed in 0.1% sodium dodecyl sulfate (SDS) to remove unadsorbed DNA. To reduce nonspecific adsorption of the labeled hybridization probe to the poly-L-lysine coated glass surface, the slides were treated with succinic anhydride. One gram of succinic anhydride was dissolved in 100 ml of 1-methyl-2-pyrrolidinone and then 100 ml of 0.2 M boric acid (pH 8.0) was added. The arrays were soaked in this solution for 10 min and then rinsed in distilled water four times for 5 min each. Immediately before use, the arrayed DNA elements were denatured by placing the slide in distilled water at 90°C for 2 min.

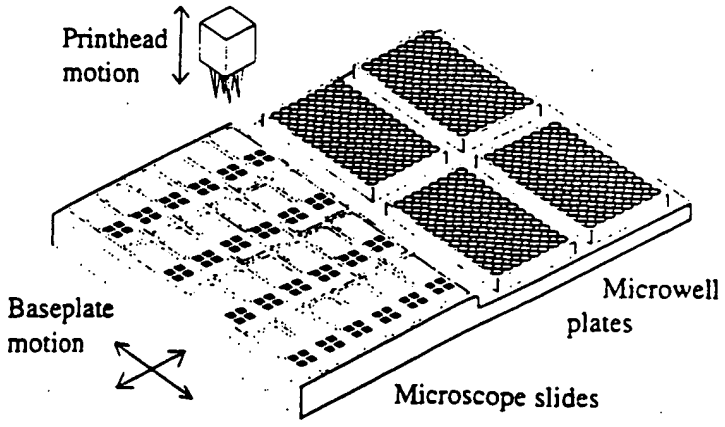


Figure 3 The layout of the arraying machine. All motions are under computer control. For more details of the arraying machine, see web page http://cmgm.stanford.edu/pbrown.

Amplification and Labeling of Hybridization Probe

The 16 chromosomes of *Saccharomyces cerevisiae* were separated using a contour-clamped homogeneous electric field (CHEF) agarose gel apparatus (Bio-Rad) (Chu et al. 1986). The 6 largest chromosomes were isolated in one gel slice and the smallest ten chromosomes in a second gel slice. The DNA from each slice was recovered using a gel extraction kit

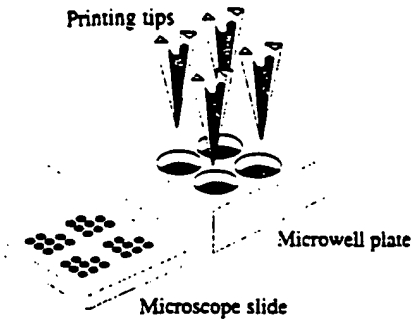


Figure 4 A close-up view of the four open-capillary printing tips. The tips are 9 mm apart and fit into four adjacent wells of a standard microwell plate and print arrays in a staggered fashion on microscope slides. For more details of the printing tips, see web page http://cmgm.stanford.edu/pbrown.

(Qiagen) and randomly amplified in a manner similar to that used in amplifying the target λ clones (Grothues et al. 1993). The main difference between this amplification procedure and the one used for the λ array elements is a filtration step between rounds A and B to remove primer-dimers and the use of a random 9-mer 3′ end on primer A. Following amplification, 2.5 μg of each of the amplified chromosome pools were separately random-primer labeled using Klenow polymerase (Amersham) with a lissamine-conjugated nucleotide analog (DuPont NEN) for the pool containing the 6 largest chromosomes and with a fluorescein-conjugated nucleotide analog (BMB) for the pool containing the smallest 10 chromosomes. The two fluorescent-labeled pools were mixed and concentrated using an ultrafiltration device (Amicon).

Hybridization

Five micrograms of the hybridization probe, consisting of both chromosome pools in 7.5 μl of TE, was denatured in a boiling water bath and then snap-cooled on ice. Concentrated hybridization solution (2.5 μl) was added to a final concentration of 5× SSC/0.1% SDS. The entire 10 μl of probe solution was transferred to the array surface, covered with a coverslip, placed in a custom-built single-slide humidity chamber, and incubated in a 60°C water bath for 12 hr. The custom-built waterproof slide chamber has a cavity just slightly bigger than a microscope slide and was kept at 100% humidity internally by the addition of 2 μl of water in a corner of the chamber. The slide was rinsed in 5× SSC/0.1% SDS for 5 min and then in 0.2× SSC/0.1% SDS for 5 min. All rinses were at room temperature. The array was then air dried, and a drop of antifade (Molecular Probes) was applied to the array under a 24-mm × 30-mm coverslip in preparation for scanning.

Detection and Analysis

A custom-built laser scanner was used to detect the two-color fluorescence hybridization signals from 1.8-cm × 1.8-cm arrays at 20-μm resolution. The glass substrate slide was mounted on a computer-controlled, two-axis translation stage (PM-500, Newport, Irvine, CA) that scanned the array over an upward-facing microscope objective (20×, 0.75NA Fluor, Nikon, Melville, NY) in a bi-directional raster pattern. A water-cooled Argon/Krypton laser (Innova 70 Spectrum, Coherent, Palo Alto, CA), operated in multiline mode, allowed for simultaneous specimen illumination at 488.0 nm and 568.2 nm. These two lines were isolated by a 488/568 dual-band excitation filter (Chroma Technology, Brattleboro, VT). An epifluorescence configuration with a dual-band 488/568 primary beam splitter (Chroma) excited both fluorophores simultaneously and directed fluorescence emissions toward the two-channel detector. Emissions were split by a secondary dichroic mirror with a 565 transition wavelength onto two multialkali cathode photomultiplier tubes (PMT; R928, Hamamatsu, Bridgewater, NJ), one with an HQ535/50 bandpass barrier filter and the other with a D630/60 bandpass barrier filter (Chroma). Preamplified PMT signals were read into a personal computer using a 12-bit analog-to-digital conversion board (RTI-834, Analog Devices, Norwood, MA), displayed in a graphics window, and stored to disk for further rendering and analysis. The back aperture of the 20× objective was deliberately underfilled by the illuminating laser beam to produce a large-diameter illuminating spot at the specimen (5-μm to 10-μm half-width). Stage scanning velocity was 100 mm/sec, and PMT signals were digitized at 100 μsec intervals. Two successive readings were summed for each pixel, such that pixel spacing in the final image was 20 μm. Beam power at the specimen was ~5 mW for each of the two lines.

The scanned image was despeckled using a graphics program (Hijaak Graphics Suite) and then analyzed using a custom image gridding program that created a spreadsheet of the average red and green hybridization intensities for each spot. The red and green hybridization intensities were corrected for optical cross talk between the fluorescein and lissamine channels, using experimentally determined coefficients.

ACKNOWLEDGMENTS

This research was supported by grant HG00450 from the National Institutes of Health–National Center for Human Genome Research, a National Science Foundation graduate fellowship to D.S., and by the Howard Hughes Medical Institute. P.O.B. is an assistant investigator of the Howard Hughes Medical Institute. We thank John Mulligan and John McCusker for help in preparing and amplifying the λ clones used in the arrays, Ren Xin Xia for writing the scanner control software and the image gridding and automatic karyotyping programs, Jeff van Ness at Darwin Molecular Corporation for suggesting the use of succinic anhydride, Stan Nelson, Linda McAllister, Joe deRisi, and Lolita Penland for helpful suggestions in the course of this work, and Joe deRisi and Linda McAllister for helpful comments on the manuscript.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Billings, P.R., C.L. Smith, and C.R. Cantor. 1991. New techniques for physical mapping of the human genome. *FASEB J.* **5:** 28–34.

Bohlander, S.K., R. Espinosa III, M.M. LeBeau, J.D. Rowley, and M.O. Diaz. 1992. A method for the rapid sequence-independent amplification of microdissected chromosomal material. *Genomics* **13:** 1322–1324.

Chu, G., D. Vollrath, and R. Davis. 1986. Separation of large DNA molecules by contour clamped homogeneous electric fields. *Science* **234:** 1582–1585.

Drmanac, R., S. Drmanac, Z. Strezoska, T. Paunesku, I. Labat, M. Zeremski, J. Snoddy, W.K. Funkhouser, B. Koop, L. Hood, et al. 1993. DNA sequence determination by hybridization: A strategy for efficient large-scale sequencing. *Science* **260:** 1649–1652.

Grothues, D., C.R. Cantor, and C.L. Smith. 1993. PCR amplification of megabase DNA with tagged random primers (T-PCR). *Nucleic Acids Res.* **21:** 1321–1322.

Hudson, T.J., L.D. Stein, S.S. Gerety, J. Ma, A.B. Castle, J. Silva, D.K. Slonim, R. Baptista, L. Kruglyak, S.H. Xu, et al. 1995. An STS-based map of the human genome. *Science* **270:** 1945–1954.

Kallioniemi, A., O.P. Kallioniemi, D. Sudar, D. Rutovitz, J.W. Gray, F. Waldman, and D. Pinkel. 1992. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* **258:** 818–821.

Maskos, U. and E.M. Southern. 1992. Parallel analysis of oligodeoxyribonucleotide (oligonucleotide) interactions. I. Analysis of factors influencing oligonucleotide duplex formation. *Nucleic Acids Res.* **20:** 1675–1678.

Nelson, S.F., J.H. McCusker, M. Sander, Y. Kee, P. Modrich, and P.O. Brown. 1995. Genomic mismatch scanning: A new approach to genetic linkage mapping. *Nature Genet.* **4:**11–17.

Pease, A.C., D. Solas, E.J. Sullivan, M.T. Cronin, C.P. Holmes, and S.P. Fodor. 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl. Acad. Sci.* **91:** 5022–5026.

Riles, L., J.E. Dutchik, A. Baktha, B.K. McCauley, E.C. Thayer, M.P. Leckie, V.V. Braden, J.E. Depke, and M.V. Olson. 1993. Physical maps of the six smallest chromosomes of *Saccharomyces cerevisiae* at a resolution of 2.6 kilobase pairs. *Genetics* **134:** 81–150.

Ross, M.T., J.D. Hoheisel, A.P. Monaco, Z. Larin, G. Zehetner, and H. Lehrach. 1992. High density gridded YAC filters: Their potential as genome mapping tools. In *Techniques for the analysis of complex genomes* (ed. Rakesh Anand), pp. 137–153. Academic Press, London, UK.

Schena, M., D. Shalon, R.W. Davis, and P.O. Brown. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270:** 467–470.

Shalon, D., 1995. "DNA micro arrays: A new tool for genetic analysis." Ph.D. thesis, Stanford University, Stanford, CA.

Zehetner, G. and H. Lehrach. 1994. The reference library system—Sharing biological material and experimental data. *Nature* **367:** 489–491.

Received March 4, 1996; accepted in revised form May 9, 1996.

Editorial office: Cold Spring Harbor Laboratory Press, 1 Bungtown Road, Cold Spring Harbor, New York 11724-2203. Phone 516-367-8492; FAX 516-367-8334.

GENOME RESEARCH (ISSN 1054-9803) is published monthly for $495 (U.S. institutional; $545 R.O.W.), $95 (individual making personal payment; $145 R.O.W., includes airlift) by Cold Spring Harbor Laboratory Press, 1 Bungtown Road, Cold Spring Harbor, New York 11724. Periodicals class postage pending is paid at Cold Spring Harbor and additional mailing offices. POSTMASTER: Send address changes to Cold Spring Harbor Laboratory Press, 10 Skyline Drive, Plainview, New York 11803-2500.

Subscriptions: Barbara Terry, Subscription Manager. Personal: U.S. $95; R.O.W. $145 (includes airlift). Institutional: U.S. $495; R.O.W. $545 (includes airlift). Orders may be sent to Cold Spring Harbor Laboratory Press, Fulfillment Department, 10 Skyline Drive, Plainview, New York 11803-2500. Telephone: Continental U.S. and Canada 1-800-843-4388; all other locations 516-349-1930. FAX 516-349-1946. Personal subscriptions must be prepaid by personal check, credit card, or money order. Claims for missing issues must be received within 4 months of issue date.

Advertising: Teresa Tiganis, Advertising Manager, Cold Spring Harbor Laboratory Press, 1 Bungtown Road, Cold Spring Harbor, New York 11724-2203. Phone: 516-367-8351; FAX 516-367-8334.

GENOME METHODS

COVER DNA microarrays for analyzing complex DNA samples. Shown is a two-color fluorescent scan of an 1.8-cm × 1.8-cm yeast array of λ clones of yeast genomic DNA. (For details, see Shalon et al., p. 639.)

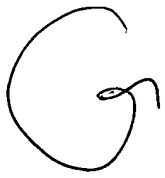
GENOME RESEARCH

Volume 6 Number 7
July 1996

RESEARCH PAPERS

LETTERS

(continued)



Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 2150–2155, March 1997
Biochemistry

Discovery and analysis of inflammatory disease-related genes using cDNA microarrays

(inflammation/human genome analysis/gene discovery)

RENU A. HELLER*†, MARK SCHENA*, ANDREW CHAI*, DARI SHALON‡, TOD BEDILION‡, JAMES GILMORE‡, DAVID E. WOOLLEY§, AND RONALD W. DAVIS*

*Department of Biochemistry, Beckman Center, Stanford University Medical Center, Stanford, CA 94305; ‡Synteni, Palo Alto, CA 94306; and §Department of Medicine, Manchester Royal Infirmary, Manchester, United Kingdom

Contributed by Ronald W. Davis, December 27, 1996

ABSTRACT cDNA microarray technology is used to profile complex diseases and discover novel disease-related genes. In inflammatory disease such as rheumatoid arthritis, expression patterns of diverse cell types contribute to the pathology. We have monitored gene expression in this disease state with a microarray of selected human genes of probable significance in inflammation as well as with genes expressed in peripheral human blood cells. Messenger RNA from cultured macrophages, chondrocyte cell lines, primary chondrocytes, and synoviocytes provided expression profiles for the selected cytokines, chemokines, DNA binding proteins, and matrix-degrading metalloproteinases. Comparisons between tissue samples of rheumatoid arthritis and inflammatory bowel disease verified the involvement of many genes and revealed novel participation of the cytokine interleukin 3, chemokine Groα and the metalloproteinase matrix metallo-elastase in both diseases. From the peripheral blood library, tissue inhibitor of metalloproteinase 1, ferritin light chain, and manganese superoxide dismutase genes were identified as expressed differentially in rheumatoid arthritis compared with inflammatory bowel disease. These results successfully demonstrate the use of the cDNA microarray system as a general approach for dissecting human diseases.

The recently described cDNA microarray or DNA-chip technology allows expression monitoring of hundreds and thousands of genes simultaneously and provides a format for identifying genes as well as changes in their activity (1, 2). Using this technology, two-color fluorescence patterns of differential gene expression in the root versus the shoot tissue of *Arabidopsis* were obtained in a specific array of 48 genes (1). In another study using a 1000 gene array from a human peripheral blood library, novel genes expressed by T cells were identified upon heat shock and protein kinase C activation (3).

The technology uses cDNA sequences or cDNA inserts of a library for PCR amplification that are arrayed on a glass slide with high speed robotics at a density of 1000 cDNA sequences per cm^2. These microarrays serve as gene targets for hybridization to cDNA probes prepared from RNA samples of cells or tissues. A two-color fluorescence labeling technique is used in the preparation of the cDNA probes such that a simultaneous hybridization but separate detection of signals provides the comparative analysis and the relative abundance of specific genes expressed (1, 2). Microarrays can be constructed from specific cDNA clones of interest, a cDNA library, or a select number of open reading frames from a genome sequencing database to allow a large-scale functional analysis of expressed sequences.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

0027-8424/97/942150-6$2.00/0
PNAS is available online at http://www.pnas.org.

Because of the wide spectrum of genes and endogenous mediators involved, the microarray technology is well suited for analyzing chronic diseases. In rheumatoid arthritis (RA), inflammation of the joint is caused by the gene products of many different cell types present in the synovium and cartilage tissues plus those infiltrating from the circulating blood. The autoimmune and inflammatory nature of the disease is a cumulative result of genetic susceptibility factors and multiple responses, paracrine and autocrine in nature, from macrophages, T cells, plasma cells, neutrophils, synovial fibroblasts, chondrocytes, etc. Growth factors, inflammatory cytokines (4), and the chemokines (5) are the important mediators of this inflammatory process. The ensuing destruction of the cartilage and bone by the invading synovial tissue includes the actions of prostaglandins and leukotrienes (6), and the matrix degrading metalloproteinases (MMPs). The MMPs are an important class of Zn-dependent metallo-endoproteinases that can collectively degrade the proteoglycan and collagen components of the connective tissue matrix (7).

This paper presents a study in which the involvement of select classes of molecules in RA was examined. Also investigated were 1000 human genes randomly selected from a peripheral human blood cell library. Their differential and quantitative expression analysis in cells of the joint tissue, in diseased RA tissue and in inflammatory bowel disease (IBD) tissues was conducted to demonstrate the utility of the microarray method to analyze complex diseases by their pattern of gene expression. Such a survey provides insight not only into the underlying cause of the pathology, but also provides the opportunity to selectively target genes for disease intervention by appropriate drug development and gene therapies.

METHODS

Microarray Design, Development, and Preparation. Two approaches for the fabrication of cDNA microarrays were used in this study. In the first approach, known human genes of probable significance in RA were identified. Regions of the clones, preferably 1 kb in length, were selected by their proximity to the 3′ end of the cDNA and for areas of least identity to related and repetitive sequences. Primers were synthesized to amplify the target regions by standard PCR protocols (3). Products were

Abbreviations: RA, rheumatoid arthritis; MMP, matrix-degrading metalloproteinase; IBD, inflammatory bowel disease; LPS, lipopolysaccharide; PMA, phorbol 12-myristate 13-acetate; TNF-α, tumor necrosis factor α; IL, interleukin; TGF-β, transforming growth factor β; GCSF, granulocyte colony-stimulating factor; MIP, macrophage inflammatory protein; MIF, migration inhibitory factor; HME, human matrix metallo-elastase; RANTES, regulated upon activation, normal T cell expressed and secreted; Gel, gelatinase; VCAM, vascular cell adhesion molecule; ICE, IL-1 converting enzyme; PUMP, putative metalloproteinase; MnSOD, manganese superoxide dismutase; TIMP, tissue inhibitor of metalloproteinase; MCP, macrophage chemotactic protein.

†To whom reprint requests should be sent at the present address: Roche Bioscience, S3–1, 3401 Hillview Avenue, Palo Alto, CA 94304.

verified by gel electrophoresis and purified with Qiaquick 96-well purification kit (Qiagen, Chatsworth, CA), lyophilized (Savant), and resuspended in 5 μl of 3× standard saline citrate (SSC) buffer for arraying. In the second approach, the microarray containing the 1056 human genes from the peripheral blood lymphocyte library was prepared as described (3).

Tissue Specimens. Rheumatoid synovial tissue was obtained from patients with late stage classic RA undergoing remedial synovectomy or arthroplasty of the knee. Synovial tissue was separated from any associated connective tissue or fat. One gram of each synovial specimen was subjected to RNA extraction within 40 min of surgical excision, or explants were cultured in serum-free medium to examine any changes under *in vitro* conditions. For IBD, specimens of macroscopically inflamed lower intestinal mucosa were obtained from patients with Crohn disease undergoing remedial surgery. The hypertrophied mucosal tissue was separated from underlying connective tissue and extracted for RNA.

Cultured Cells. The Mono Mac-6 (MM6) monocytic cells (8) were grown in RPMI medium. Human chondrosarcoma SW1353 cells, primary human chondrocytes, and synoviocytes (9, 10) were cultured in DMEM; all culture media were supplemented with 10% fetal bovine serum, 100 μg/ml streptomycin, and 500 units/ml penicillin. Treatment of cells with lipopolysaccharide (LPS) endotoxin at 30 ng/ml, phorbol 12-myristate 13-acetate (PMA) at 50 ng/ml, tumor necrosis factor α (TNF-α) at 50 ng/ml, interleukin (IL)-1β at 30 ng/ml, or transforming growth factor-β (TGF-β) at 100 ng/ml is described in the figure legends.

Fluorescent Probe, Hybridization, and Scanning. Isolation of mRNA, probe preparation, and quantitation with *Arabidopsis* control mRNAs was essentially as described (3) except for the following minor modification. Following the reverse transcriptase step, the appropriate Cy3- and Cy5-labeled samples were pooled: mRNA degraded by heating the sample to 65°C for 10 min with the addition of 5 μl of 0.5M NaOH plus 0.5 ml of 10 mM EDTA. The pooled cDNA was purified from unincorporated nucleotides by gel filtration in Centri-spin columns (Princeton Separations, Adelphia, NJ). Samples were lyophilized and dissolved in 6 μl of hybridization buffer (5× SSC plus 0.2% SDS). Hybridizations, washes, scanning, quantitation procedures, and pseudocolor representations of fluorescent images have been described (3). Scans for the two fluorescent probes were normalized either to the fluorescence intensity of *Arabidopsis* mRNAs spiked into the labeling reactions (see Figs. 2–4) or to the signal intensity of β-actin and glyceraldehyde-3-phosphate dehydrogenase (GAPDH; see Fig. 5).

RESULTS

Ninety-Six-Gene Microarray Design. The actions of cytokines, growth factors, chemokines, transcription factors, MMPs, prostaglandins, and leukotrienes are well recognized in inflammatory disease, particularly RA (11–14). Fig. 1 displays the selected genes for this study and also includes control cDNAs of housekeeping genes such as β-actin and GAPDH and genes from *Arabidopsis* for signal normalization and quantitation (row A, columns 1–12).

Defining Microarray Assay Conditions. Different lengths and concentrations of target DNA were tested by arraying PCR-

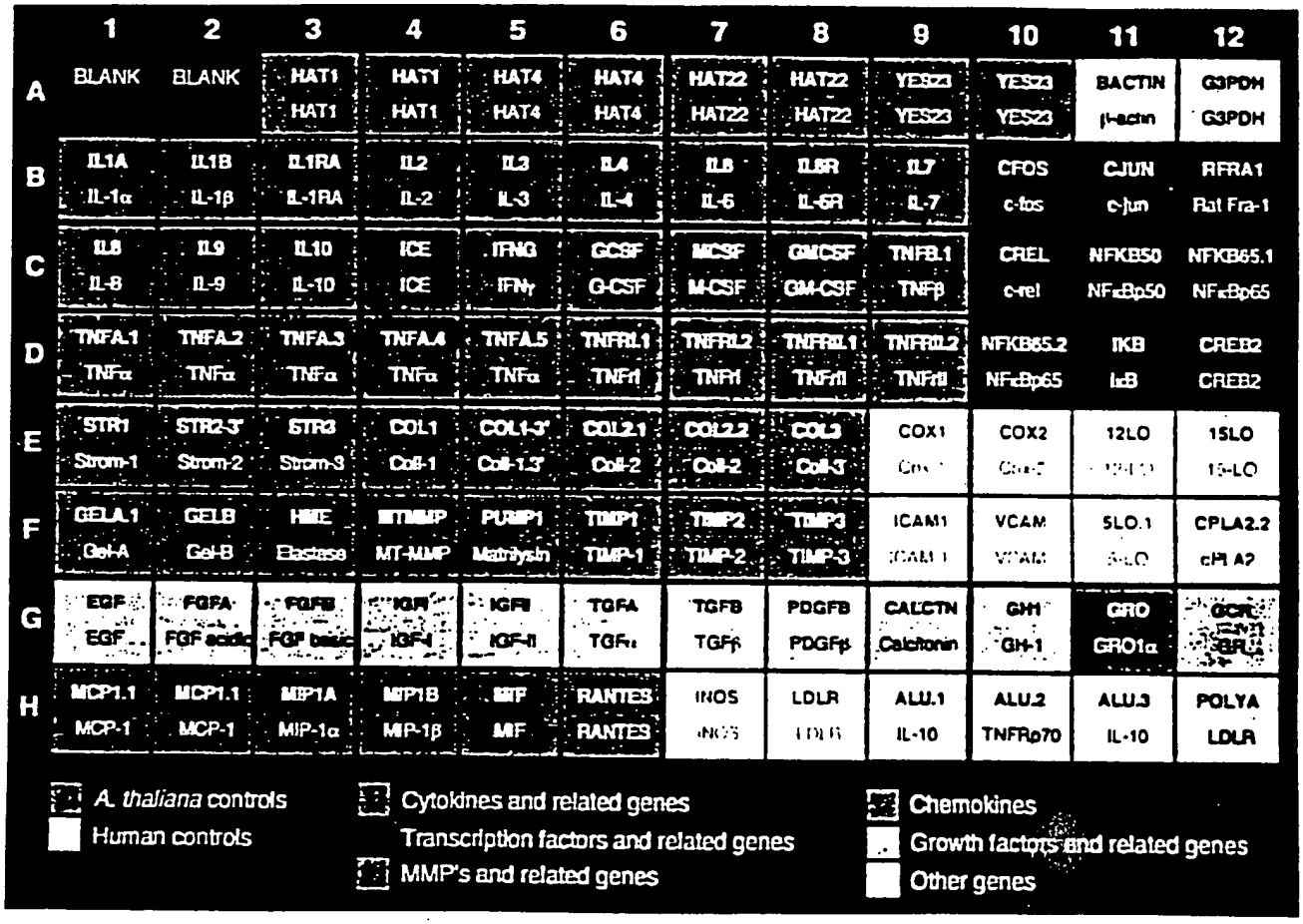


FIG. 1. Ninety-six-element microarray design. The target element name and the corresponding gene are shown in the layout. Some genes have more than one target element to guarantee specificity of signal. For TNF the targets represent decreasing lengths of 1, 0.8, 0.6, 0.4, and 0.2 kb from left to right.

amplified products ranging from 0.2 to 1.2 kb at concentrations of 1 $\mu g/\mu l$ or less. No significant difference in the signal levels was observed within this range of target size and only with 0.2-kb length was a signal reduced upon an 8-fold dilution of the 1 $\mu g/\mu l$ sample (data not shown). In this study the average length of the targets was 1 kb, with a few exceptions in the range of ~300 bp, arrayed at a concentration of 1 $\mu g/\mu l$. Normally one PCR provided sufficient material to fabricate up to 1000 microarray targets.

In considering positional effects in the development of the targets for the microarrays, selection was biased toward the 3′ proximal regions, because the signal was reduced if the target fragment was biased toward the 5′ end (data not shown). This result was anticipated since the hybridizing probe is prepared by reverse transcription with oligo(dT)-primed mRNA and is richer in 3′ proximal sequences. Cross-hybridizations of probes to targets of a gene family were analyzed with the matrix metalloproteinases as the example because they can show regions of sequence identities of greater than 70%. With collagenase-1 (Col-1) and collagenase-2 (Col-2) genes as targets with up to 70% sequence identity, and stromelysin-1 (Strom-1) and stromelysin-2 (Strom-2) genes with different degrees of identity, our results showed that a short region of overlap, even with 70–90% sequence identity, produced a low level of cross-hybridization. However, shorter regions of identity spread over the length of the target resulted in cross-hybridization (data not shown). For closely related genes, targets were designed by avoiding long stretches of homology. For members of a gene family two or more target regions were included to discriminate between specificity of signal versus cross-hybridization.

Monitoring Differential Expression in Cultured Cell Lines. In RA tissue, the monocyte/macrophage population plays a prominent role in phagocytic and immunomodulatory activities. Typ-

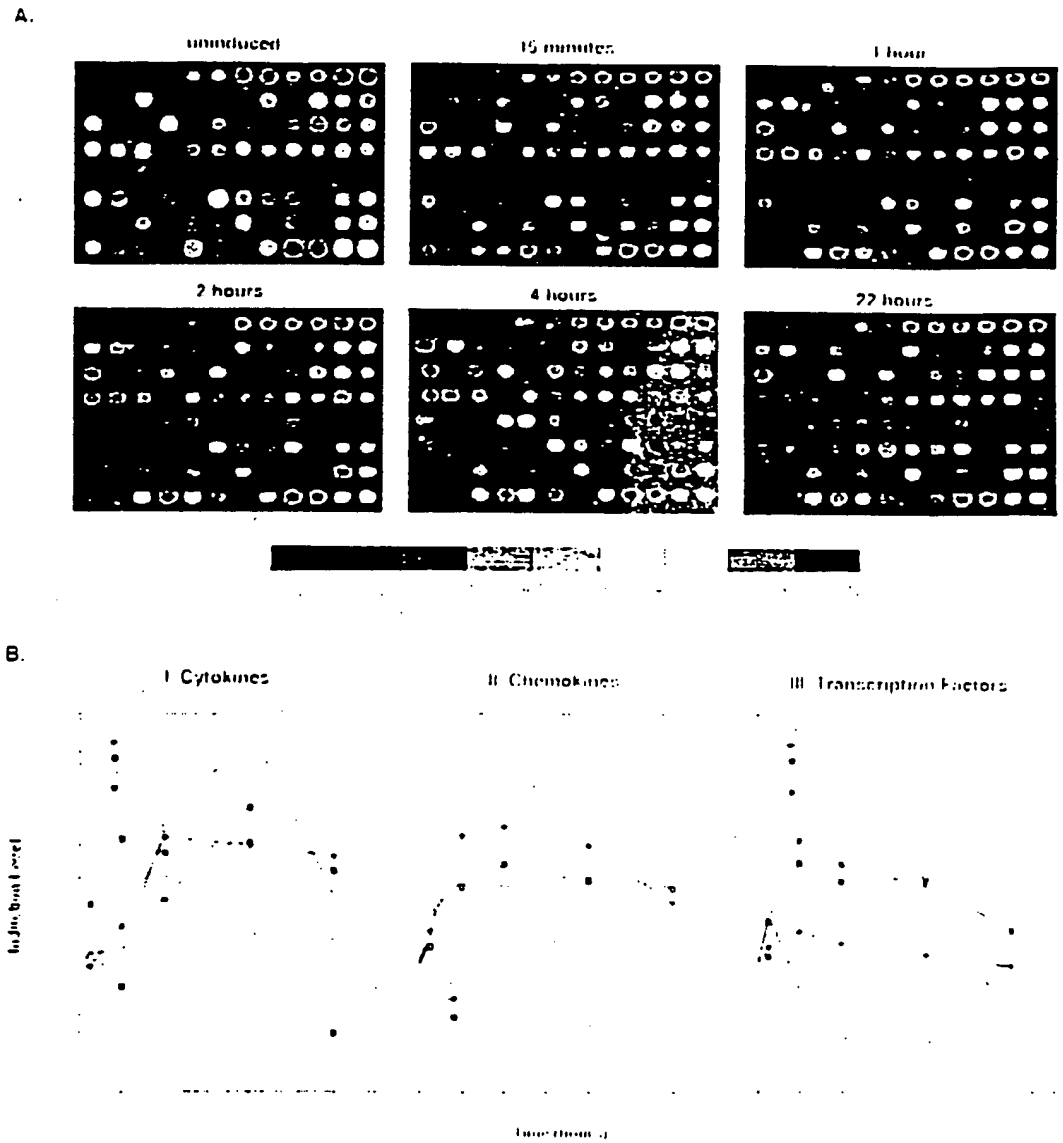


FIG. 2. Time course for LPS/PMA-induced MM6 cells. Array elements are described in Fig. 1. (*A*) Pseudocolor representations of fluorescent scans correspond to gene expression levels at each time point. The array is made up of 8 *Arabidopsis* control targets and 86 human cDNA targets, the majority of which are genes with known or suspected involvement in inflammation. The color bars provide a comparative calibration scale between arrays and are derived from the *Arabidopsis* mRNA samples that are introduced in equal amounts during probe preparation. Fluorescent probes were made by labeling mRNA from untreated MM6 cells or LPS and PMA treated cells. mRNA was isolated at indicated times after induction. (*B I–III*) The two-color samples were cohybridized, and microarray scans provided the data for the levels of select transcripts at different time points relative to abundance at time zero. The analysis was performed using normalized data collected from 8-bit images.

ically these cells, when triggered by an immunogen, produce the proinflammatroy cytokines TNF and IL-1. We have used the monocyte cell line MM6 and monitored changes in gene expression upon activation with LPS endotoxin, a component of Gram-negative bacterial membranes, and PMA, which augments the action of LPS on TNF production (15). RNA was isolated at different times after induction and used for cDNA probe preparation. From this time course it was clear that TNF expression was induced within 15 min of treatment, reached maximum levels in 1 hr, remained high until 4 hr and subsequently declined (Fig. 2*A*). Many other cytokine genes were also transiently activated, such as IL-1α and -β, IL-6, and granulocyte colony-stimulating factor (GCSF). Prominent chemokines activated were IL-8, macrophage inflammatory protein (MIP)-1β, more so than MIP-1α, and Groα or melanoma growth stimulatory factor. Migration inhibitory factor (MIF) expressed in the uninduced state declined in LPS-activated cells. Of the immediate early genes, the noticeable ones were *c-fos*, *fra-1*, *c-jun*, NF-κBp50, and IκB, with *c-rel* expression observed even in the uninduced state (Fig. 2*B*). These expression patterns are consistent with reported patterns of activation of certain LPS- and PMA-induced genes (12). Demonstrated here is the unique ability of this system to allow parallel visualization of a large number of gene activities over a period of time.

SW1353 cells is a line derived from malignant tumors of the cartilage and behaves much like the chondrocytes upon stimulation with TNF and IL-1 in the expression of MMPs (9). In addition to confirming our earlier observations with Northern blots on Strom-1, Col-1, and Col-3 expression (9), gelatinase (Gel) A, putative metalloproteinase (PUMP)-1 membrane-type matrix metalloproteinase, tissue inhibitors of matrix metalloproteinases or tissue inhibitor of metalloproteinase 1 (TIMP-1), -2, and -3 were also expressed by these cells together with the human matrix metallo-elastase (HME; Fig. 3*A*). HME induction was estimated to be ≈50-fold and was greater than any of the other MMPs examined (Fig. 3*B*). This result was unexpected because HME is reportedly expressed only by alveolar macrophage and placental cells (16). Expression of the cytokines and chemokines, IL-6, IL-8, MIF, and MIP-1β was also noted. A variety of other genes, including certain transcription factors, were also up-regulated (Fig. 3), but the overall time-dependent expression of genes in the SW1353 cells was qualitatively distinct from the MM6 cells.

Quantitation of differential gene expression (Figs. 2*B* and 3*B*) was achieved with the simultaneous hybridization of Cy3-labeled cDNA from untreated cells and Cy5-labeled cDNA from treated samples. The estimated increases in expression from these microarrays for a select number of genes including IL-1β, IL-8, MIP-1β, TNF, HME, Col-1, Col-3, Strom-1, and Strom-2 were compared with data collected from dot blot analysis. Results (not shown) were in close agreement and confirmed our earlier observations on the use of the microarray method for the quantitation of gene expression (3).

Expression Profiles in Primary Chondrocytes and Synoviocytes of Human RA Tissue. Given the sensitivity and the specificity of this method, expression profiles of primary synoviocytes and chondrocytes from diseased tissue were examined. Without prior exposure to inducing agents, low level expression of c-*jun*, GCSF, IL-3, TNF-β, MIF, and RANTES (regulated upon activation, normal T cell expressed and secreted) was seen as well as expression of MMPs, GelA, Strom-1, Col-1, and the three TIMPs. In this case, Col-2 hybridization was considered to be nonspecific because the second Col-2 target taken from the 3′ end of the gene gave no

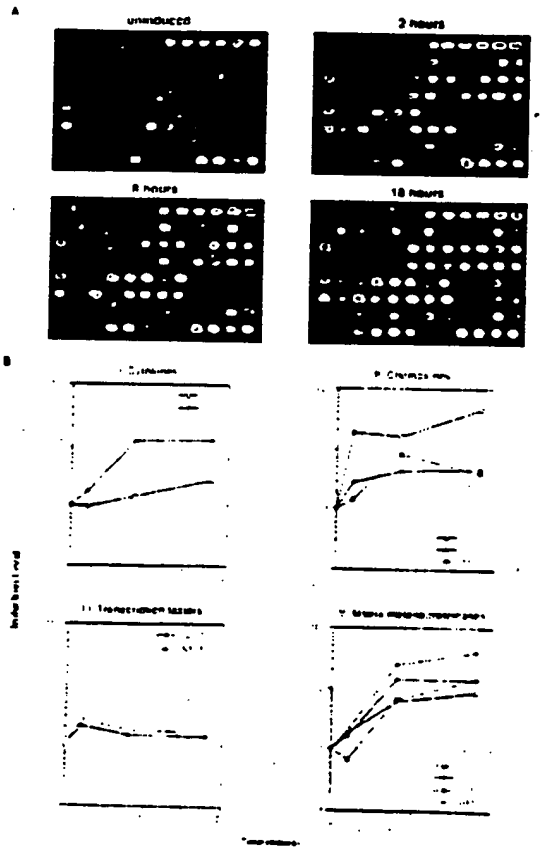


FIG. 3. Time course for IL-1β and TNF-induced SW1353 cells using the inflammation array (Fig. 1). (*A*) Pseudocolor representation of fluorescent scans correspond to gene expression levels at each time point. (*B I–IV*) Relative levels of selected genes at different time points compared with time zero.

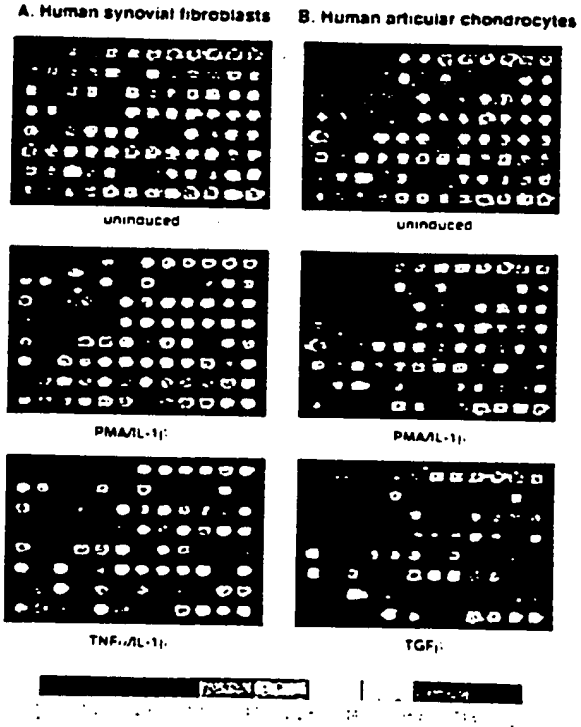


FIG. 4. Expression profiles for early passage primary synoviocytes and chondrocytes isolated from RA tissue, cultured in the presence of 10% fetal calf serum and activated with PMA and IL-1β, or TNF and IL-1β, or TGF-β for 18 hr. The color bars provide a comparative calibration scale between arrays and are derived from the *Arabidopsis* mRNA samples that are introduced in equal amounts during probe preparation

signal. Treatment more so with PMA and IL-1, than TNF and IL-1, produced a dramatic up-regulation in expression of several genes in both of these primary cell types. These genes are as follows: the cytokine IL-6, the chemokines IL-8 and Gro-1α, and the MMPs; Strom-1, Col-1, Col-3, and HME: and the adhesion molecule, vascular cell adhesion molecule 1 (VCAM-1). The surprise again is HME expression in these primary cells, for reasons discussed above. From these results, the expression profiles of synoviocytes and the chondrocytes appear very similar; the differences are more quantitative than qualitative. Treatment of the primary chondrocytes with the anabolic growth factor TGF-β had an interesting profile in that it produced a remarkable down-regulation of genes expressed in both the untreated and induced state (Fig. 4).

Given the demonstrated effectiveness of this technology, a comparative analysis of two different inflammatory disease states was conducted with probes made from RA tissue and IBD samples. RA samples were from late stage rheumatoid synovial tissue, and IBD specimens were obtained from inflamed lower intestinal mucosa of patients with Crohn disease. With both the 96-element known gene microarray and the 1000-gene microarray of cDNAs selected from a peripheral human blood cell library (3), distinct differences in gene expression patterns were evident. On the 96-gene array, RA tissue samples from different affected individuals gave similar profiles (data not shown) as did different samples from the same individual (Fig. 5). These patterns were notably similar to those observed with primary synoviocytes and chondrocytes (Fig. 4). Included in the list of prominently up-regulated genes are IL-6, the MMPs Strom-1, Col-1, GelA, HME, and in

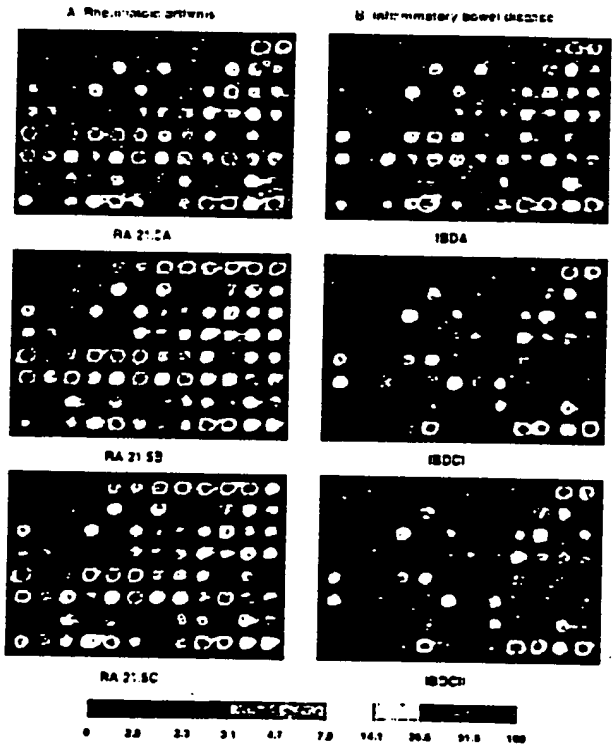


FIG. 5. Expression profiles of RA tissue (*A*) and IBD tissue (*B*). mRNA from RA tissue samples obtained from the same individual was isolated directly after excision (RA 21.5A) or maintained in culture without serum for 2 hr (RA 21.5B) or for 6 hr (RA 21.5C). Profiles from tissue samples of two other individuals (data not shown) were remarkably similar to the ones shown here. IBD-A and IBD-CI are from mRNA samples prepared directly after surgery from two separate individuals. For the IBD-CII probe, the tissue sample was cultured in medium without serum for 2 hr before mRNA preparation.

certain samples PUMP, TIMPs, particularly TIMP-1 and TIMP-3, and the adhesion molecule VCAM. Discernible levels of macrophage chemotactic protein 1 (MCP-1), MIF and RANTES were also noted. IBD samples were in comparison, rather subdued although IL-1 converting enzyme (ICE), TIMP-1, and MIF were notable in all the three different IBD samples examined here. In IBD-A, one of three individual samples, ICE, VCAM, Groα, and MMP expression was more pronounced than in the others.

We also made use of a peripheral blood cDNA library (3) to identify genes expressed by lymphocytes infiltrating the inflamed tissues from the circulating blood. With the 1046-element array of randomly selected cDNAs from this library, probes made from RA and IBD samples showed hybridizations to a large number of genes. Of these, many were common between the two disease tissues while others were differentially expressed (data not shown). A complete survey of these genes was beyond the scope of this study, but for this report we picked three genes that were up-regulated in the RA tissue relative to IBD. These cDNAs were sequenced and identified by comparison to the GenBank database. They are TIMP-1, apoferritin light chain, and manganese superoxide dismutase (MnSOD). Differential expression of MnSOD was only observed in samples of RA tissue explants maintained in growth medium without serum for anywhere between 2 to 16 hr. These results also indicate that the expression profile of genes can be altered when explants are transferred to culture conditions.

DISCUSSION

The speed, ease, and feasibility of simultaneously monitoring differential expression of hundreds of genes with the cDNA microarray based system (1–3) is demonstrated here in the analysis of a complex disease such as RA. Many different cell types in the RA tissue; macrophages, lymphocytes, plasma cells, neutrophils, synoviocytes, chondrocytes, etc. are known to contribute to the development of the disease with the expression of gene products known to be proinflammatory. They include the cytokines, chemokines, growth factors, MMPs, eicosanoids, and others (7, 11–14), and the design of the 96-element known gene microarray was based on this knowledge and depended on the availability of the genes. The technology was validated by confirming earlier observations on the expression of TNF by the monocyte cell line MM6, and of Col-1 and Col-3 expression in the chondrosarcoma cells and articular chondrocytes (9, 12). In our time-dependent survey the chronological order of gene activities in and between gene families was compared and the results have provided unprecedented profiles of the cytokines (TNF, IL-1, IL-6, GCSF, and MIF), chemokines (MIP-1α, MIP-1β, IL-8, and Gro-1), certain transcription factors, and the matrix metalloproteinases (GelA, Strom-1, Col-1, Col-3, HME) in the macrophage cell line MM6 and in the SW1353 chondrosarcoma cells.

Earlier reports of cytokine production in the diseased state had established a model in which TNF is a major participant in RA. Its expression reportedly preceded that of the other cytokines and effector molecules (4). Our results strongly support these results as demonstrated in the time course of the MM6 cells where TNF induction preceded that of IL-1α and IL-β followed by IL-6 and GCSF. These expression profiles demonstrate the utility of the microarrays in determining the hierarchy of signaling events.

In the SW1353 chondrosarcoma cells, all the known MMPs and TIMPs were examined simultaneously. HME expression was discovered, which previously had been observed in only the stromal cells and alveolar macrophages of smoker's lungs and in placental tissue. Its presence in cells of the RA tissue is meaningful because its activity can cause significant destruction of elastin and basement membrane components (16, 17). Expression profiles of synovial fibroblasts and articular chondrocytes were remarkably similar and not too different from the SW1353 cells, indicating that the fibroblast and the chondrocyte can play equally aggressive roles in joint erosion. Prominent genes expressed were

the MMPs, but chemokines and cytokines were also produced by these cells. The effect of the anabolic growth factor TGF-β was profoundly evident in demonstrating the down regulation of these catabolic activities.

RA tissue samples undeniably reflected profiles similar to the cell types examined. Active genes observed were IL-3, IL-6, ICE, the MMPs including HME and TIMPs, chemokines IL-8, Groα, MIP, MIF, and RANTES, and the adhesion molecule VCAM. Of the growth factors, fibroblast growth factor β was observed most frequently. In comparison, the expression patterns in the other inflammatory state (i.e., IBD) were not as marked as in the RA samples, at least as obtained from the tissue samples selected for this study.

As an alternative approach, the 1046 cDNA microarray of randomly selected genes from a lymphocyte library was used to identify genes expressed in RA tissue (3). Many genes on this array hybridized with probes made from both RA and IBD tissue samples. The results are not surprising because inflammatory tissue is abundantly supplied with cell types infiltrating from the circulating blood, made apparent also by the high levels of chemokine expression in RA tissue. Because of the magnitude of the effort required to identify all the hybridized genes, we have for this report chosen to describe only three differentially expressed genes mainly to verify this method of analysis.

Of the large number of genes observed here, a fair number were already known as active participants in inflammatory disease. These are TNF, IL-1, IL-6, IL-8, GCSF, RANTES, and VCAM. The novel participants not previously reported are HME, IL-3, ICE, and Groα. With our discovery of HME expression in RA, this gene becomes a target for drug intervention. ICE is a cysteine protease well known for its IL-1β processing activity (18), and recognized for its role in apoptotic cell death (19). Its expression in RA tissue is intriguing. IL-3 is recognized for its growth-promoting activity in hematopoietic cell lineages, is a product of activated T cells (20), and its expression in synoviocytes and chondrocytes of RA tissue is a novel observation.

Like IL-8, Groα, is a C-X-C subgroup chemokine and is a potent neutrophil and basophil chemoattractant. It down-regulates the expression of types I and III interstitial collagens (21, 22) and is seen here produced by the MM6 cells, in primary synoviocytes, and in RA tissue. With the presence of RANTES, MCP, and MIP-1β, the C-C chemokines (23) migration and infiltration of monocytes, particularly T cells, into the tissue is also enhanced (5) and aid in the trafficking and recruitment of leukocytes into the RA tissue. Their activation, phagocytosis, degranulation, and respiratory bursts could be responsible for the induction of MnSOD in RA. MnSOD is also induced by TNF and IL-1 and serves a protective function against oxidative damage. The induction of the ferritin light chain encoding gene in this tissue may be for reasons similar to those for MnSOD. Ferritin is the major intracellular iron storage protein and it is responsive to intracellular oxidative stress and reactive oxygen intermediates generated during inflammation (24, 25). The active expression of TIMP-1 in RA tissue, as detected by the 1000-element array, is no surprise because our results have repeatedly shown TIMP-1 to be expressed in the constitutive and induced states of RA cells and tissues.

The suitability of the cDNA microarray technology for profiling diseases and for identifying disease related genes is well documented here. This technology could provide new targets for drug development and disease therapies, and in doing so allow for improved treatment of chronic diseases that are challenging because of their complexity.

We would like to thank the following individuals for their help in obtaining reagents or providing cDNA clones to use as templates in target preparation: N. Arai, P. Cannon, D. R. Cohen, T. Curran, V. Dixit, D. A. Geller, G. I. Goldberg, M. Karin, M. Lotz, L. Matrisian, G. Nolan, C. Lopez-Otin, T. Schall, S. Shapiro, I. Verma, and H. Van Wart. Support for R.W.D., M.S., and R.A.H. was provided by the National Institutes of Health (Grants R37HG00198 and HG00205).

1. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. (1995) *Science* 270, 467–470.
2. Shalon, D., Smith, S. & Brown, P. O. (1996) *Genome Res.* 6, 639–645.
3. Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O. & Davis, R. W. (1996) *Proc. Natl. Acad. Sci. USA* 93, 10614–10619.
4. Feldmann, M., Brennan F. M. & Maini, R. N. (1996) *Rheumatoid Arthritis Cell* 85, 307–310.
5. Schall, T. J. (1994) in *The Cytokine Handbook*, ed. Thomson, A. W. (Academic, New York), 2nd Ed., pp. 410–460.
6. Lotz, M. F., Bianco, J., Von Kempis, J., Dudler, J., Maier, R., Villiger P. M. & Geng, Y. (1995) *J. Rheumatol.* 22, Supplement 43, 104–108.
7. Birkedal-Hansen, H., Moore, W. G. I., Bodden, M. K., Windsor, L. J., Birkedal-Hansen, B., DeCarlo, A. &. Engler, J. A. (1993) *Crit. Rev. Oral Biol. Med.* 4, 197–250.
8. Zeigler-Heitbrock, H. W. L., Thiel, E., Futterer, A., Volker, H., Wirtz, A. & Reithmuller, G. (1988) *Int. J. Cancer* 41, 456–461.
9. Borden, P., Solymar, D., Sucharczuk, A., Lindman, B., Cannon, P. & Heller, R. A. (1996) *J. Biol. Chem.* 271, 23577–23581.
10. Gadher, S. J. & Woolley, D. E. (1987) *Rheumatol. Int.* 7, 13–22.
11. Harris, E. D., Jr. (1990) *New Engl. J. Med.* 322, 1277–1289.
12. Firestein, G. S. (1996) in *Textbook of Rheumatology*, eds. Kelly, W. N., Harris, E. D., Ruddy, S. & Sledge, C. B. (Saunders, Philadelphia), 5th Ed. pp. 5001–5047.
13. Alvaro-Garcia, J. M., Zvaifler, Nathan J., Brown, C. B., Kaushansky, K. & Firestein, Gary S. (1991) *J. Immunol.* 146, 3365–3371.
14. Firestein, G. S., Alvaro-Grarcia, J. M. & Maki, R. (1990) *J. Immunol.* 144, 3347–3352.
15. Pradines-Figueres, A. & Raetz, C. R. H. (1992) *J. Biol. Chem.* 267, 23261–23268.
16. Shapiro, S. D., Kobayashi, D. L. & Ley, T. J. (1993) *J. Biol. Chem.* 208, 23824–23829.
17. Shipley, M. J., Wesselschmidt, R. L., Kobayashi, D. K., Ley, T. J. & Shapiro, S. D. (1996) *Proc. Natl. Acad. Sci. USA* 93, 3042–3946.
18. Cerretti, D. P., Kozlosky, C. J., Mosley, B., Nelson, N., Van Ness, K., Greenstreet, T. A., March, C. J., Kronheim, S. R., Druck, T., Cannizaro, L. A., Huebner, K. & Black, R. A. (1992) *Science* 256, 97–100.
19. Miura, M., Zhu, H., Rotello, R., Hartweig, E. A. & Yuan, J. (1993) *Cell* 75, 653–660.
20. Arai, K., Lee, F., Miyajima, A., Shoichiro, M., Arai, N. & Takashi, Y. (1990) *Annu. Rev. Biochem.* 59, 783–836.
21. Geiser, T., Dewald, B., Ehrengruber, M. U., Lewis, I. C. & Baggiolini, M. (1993) *J. Biol. Chem.* 268, 15419–15424.
22. Unemori, E. N., Amento, E. P., Bauer, E. A. & Horuk, R. (1993) *J. Biol. Chem.* 268, 1338–1342.
23. Robinson, E., Keystone, E. C., Schall, T. J., Gillet, N. & Fish, E. N. (1995) *Clin. Exp. Immunol.* 101, 398–407.
24. Roeser, H. (1980) in *Iron Metabolism in Biochemistry and Medicine*, eds. Jacobs, A. & Worwood, M. (Academic, New York), Vol. 2, pp. 605–640.
25. Kwak, E. L., Larochelle, D. A., Beaumont, C., Torti, S. V. & Torti, F. M. (1995) *J. Biol. Chem.* 270, 15285–15293.